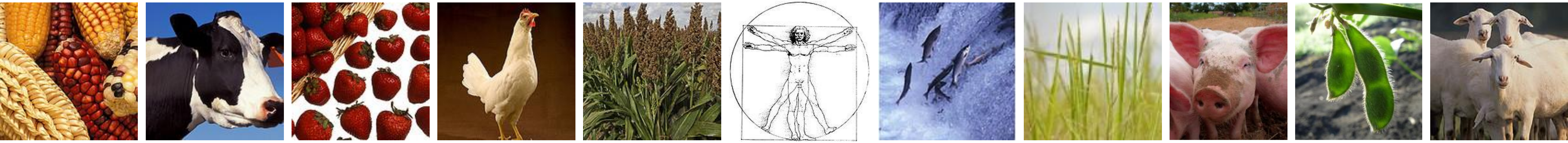


NCBI RefSeq Resources for Plant Genomics

Anjana R Vatsan(raina@nih.gov)

Functional Genomics Workshop

PAG XXVIII January 13, 2020

NIH
U.S. National Library of Medicine
National Center for Biotechnology Information

Also from NCBI!

Visit NCBI Booth **321** Contact us **info@ncbi.nlm.nih.gov**

Day	Time	Topic
Monday	12:50 pm – 3:00 pm *Pacific Salon 1*	**NCBI Genome Resources Workshop**
Tuesday	11:10 am *California*	**NCBI BLAST: Enhanced Web Usability through New Result Page and Effective Genomic Data Access** *Digital Tools and Resources Session 3*
Wednesday	11:50 am *California*	**Federated Cloud Access to Datasets through Indexing and/or Graphs!** *Digital Tools and Resources Session 4*

NCBI Genome Resources Workshop

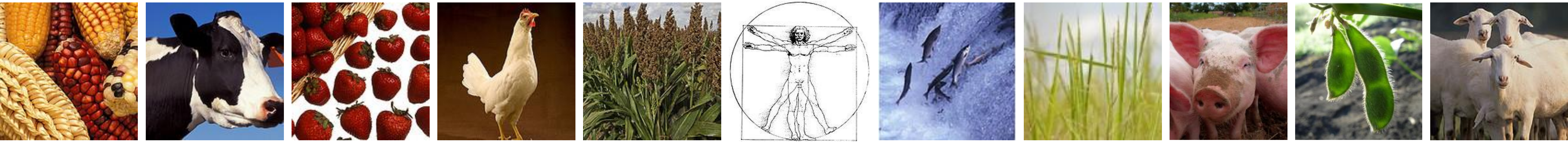
Monday January 13, 2020, 12:50 – 3:00 pm, Pacific Salon 1

Time	Topic
12:55 – 1:15	NCBI Wants Your Sequence Data! How Do I Get It There? *Ilene Mizrachi*
1:15 – 1:35	Annotation of Eukaryote Genomes at NCBI *Jinna Hoffman*
1:35 – 1:55	Accessing Homologous Gene Datasets at NCBI *Nuala O'Leary*
1:55 – 2:15	The New PubMed Is Here! *Kathi Canese*
2:15 – 2:35	Taxonomy Lookup; Data Retrieval: How to Find and Stream Genomic Data in the Cloud! *Ben Busby*

Visit NCBI Booth **321** Contact us **info@ncbi.nlm.nih.gov**

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Watch NCBI News for updates!
http://www.ncbi.nlm.nih.gov/news/
https://www.youtube.com/user/NCBINLM



NCBI RefSeq Resources for Plant Genomics

Anjana R Vatsan(raina@nih.gov)

Functional Genomics Workshop

PAG XXVIII January 13, 2020

U.S. National Library of Medicine
National Center for Biotechnology Information

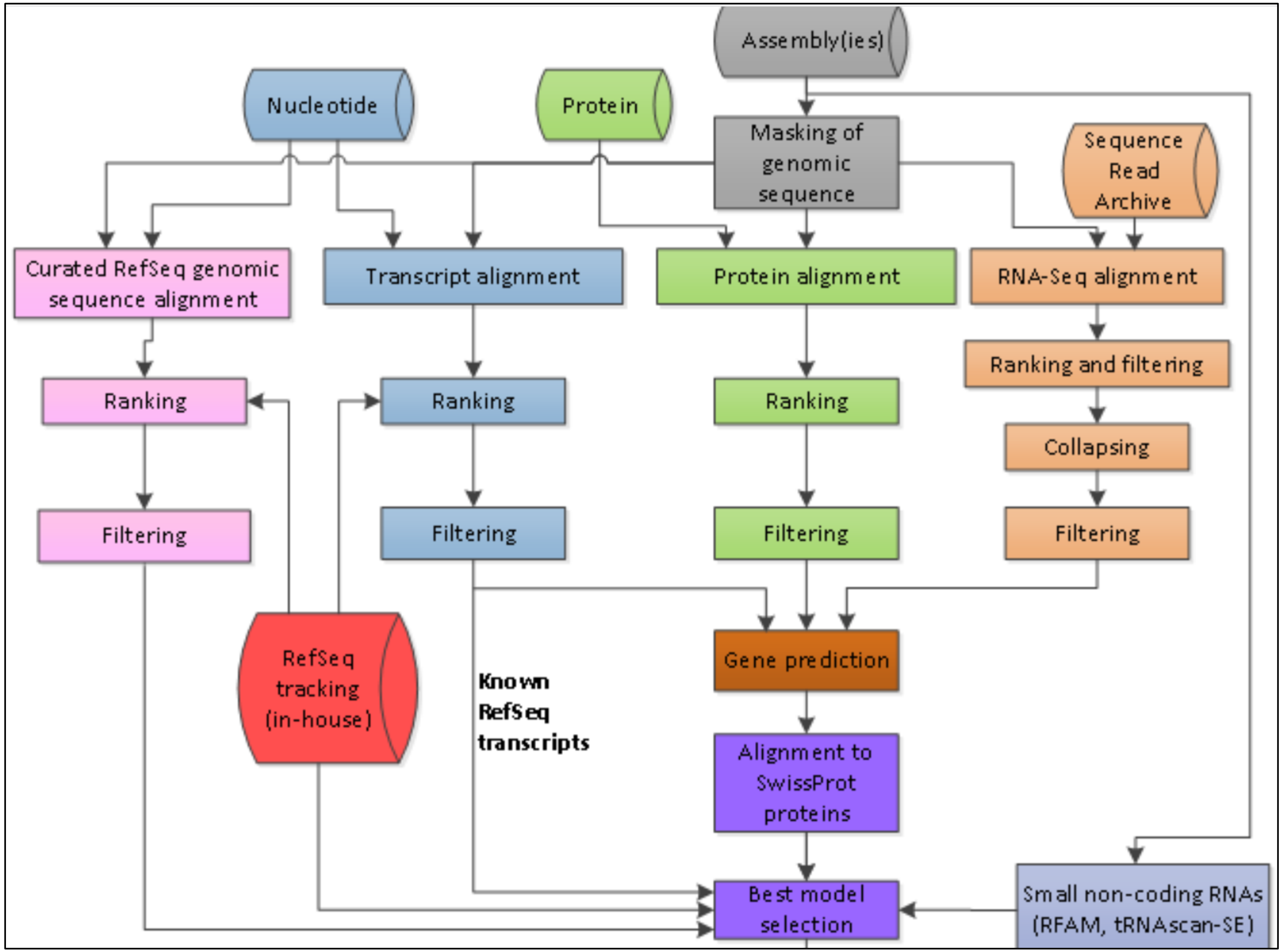
RefSeq: NCBI Reference Sequence Database

RefSeq is a public database of nucleotide and protein sequences which are derived, in most part, from genome assemblies that are submitted to International Nucleotide Sequence Database Collaboration (INSDC), by one of the following methods.

- Computationally using the Genome Annotation Pipeline
- Manual curation
- Propagation from model organism databases
  e.g. Arabidopsis thaliana

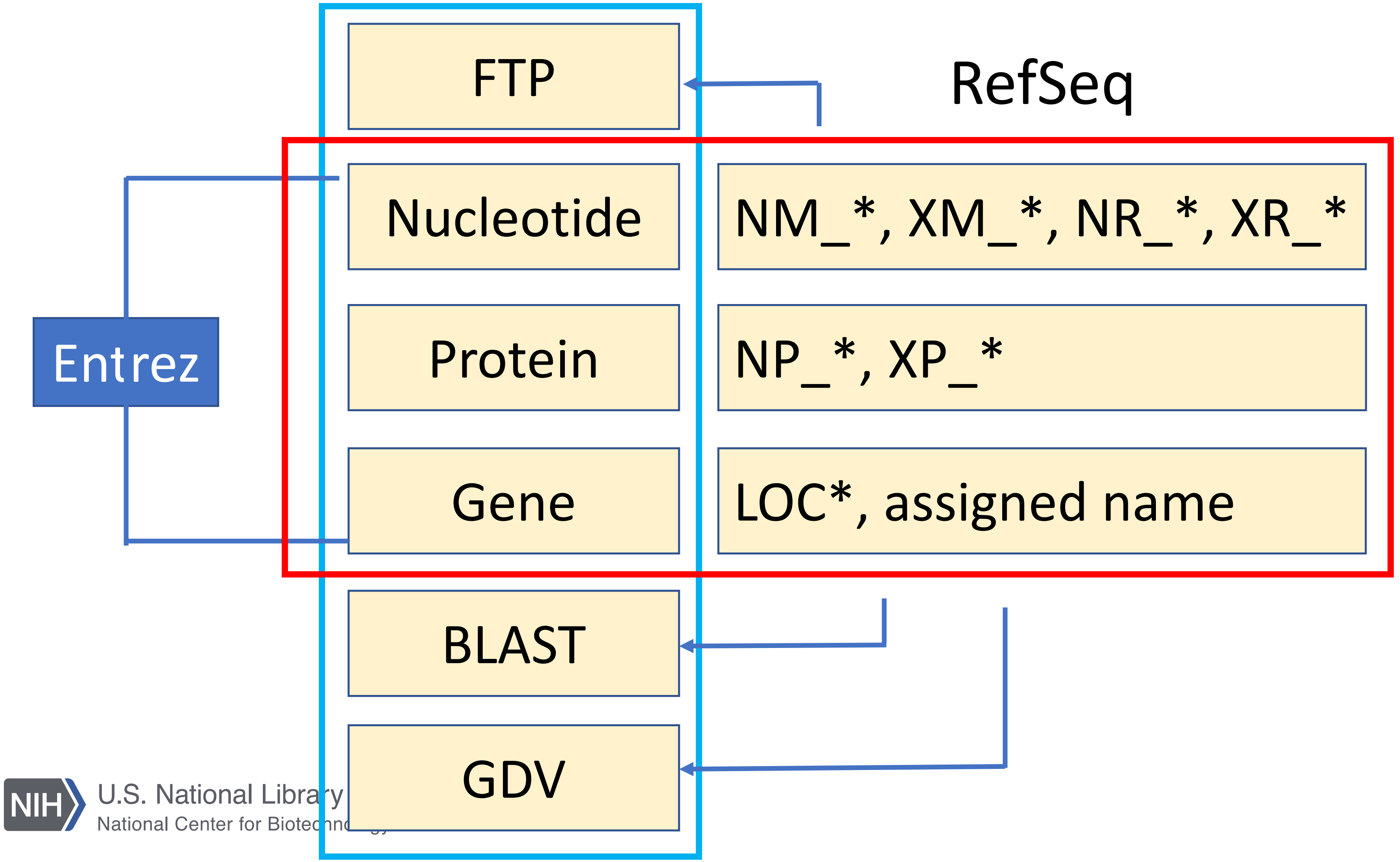
https://www.ncbi.nlm.nih.gov/refseq/

Annotation Pipeline



Annotation Output

ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/



FTP

RefSeq

Entrez

Nucleotide

NM_*, XM_*, NR_*, XR_*

Protein

NP_*, XP_*

Gene

LOC*, assigned name

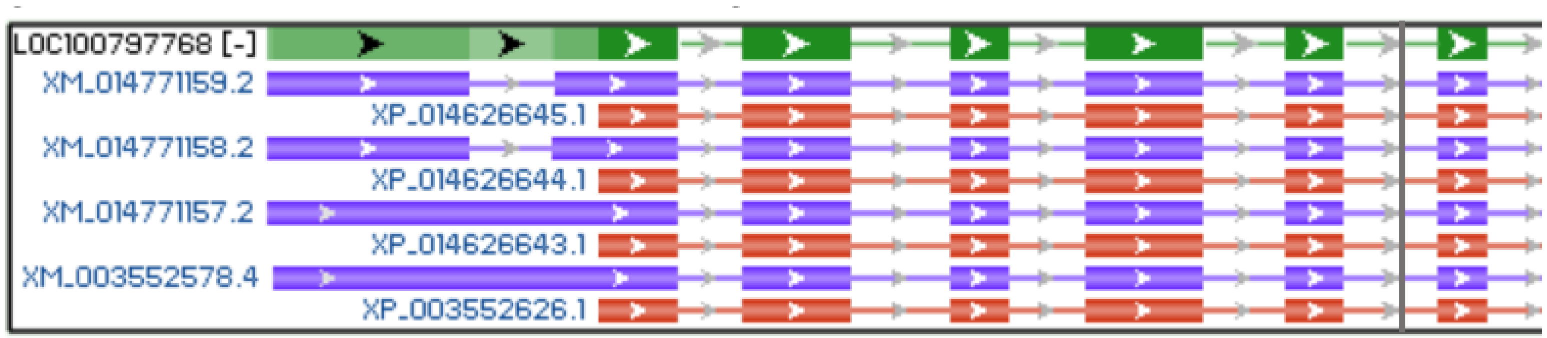
BLAST

GDV

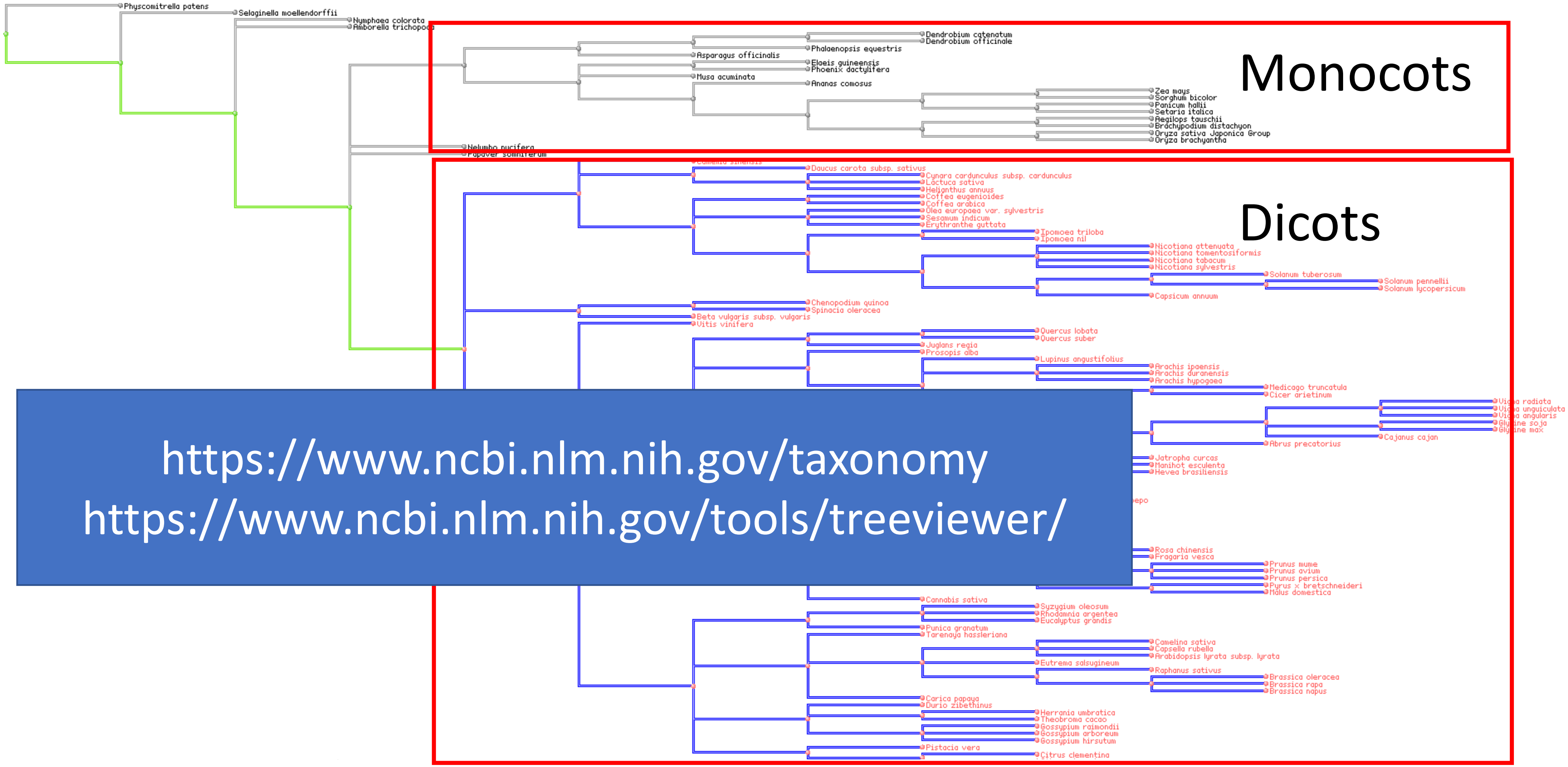
RefSeq Data organization

Gene

Alternatively spliced transcripts and corresponding proteins

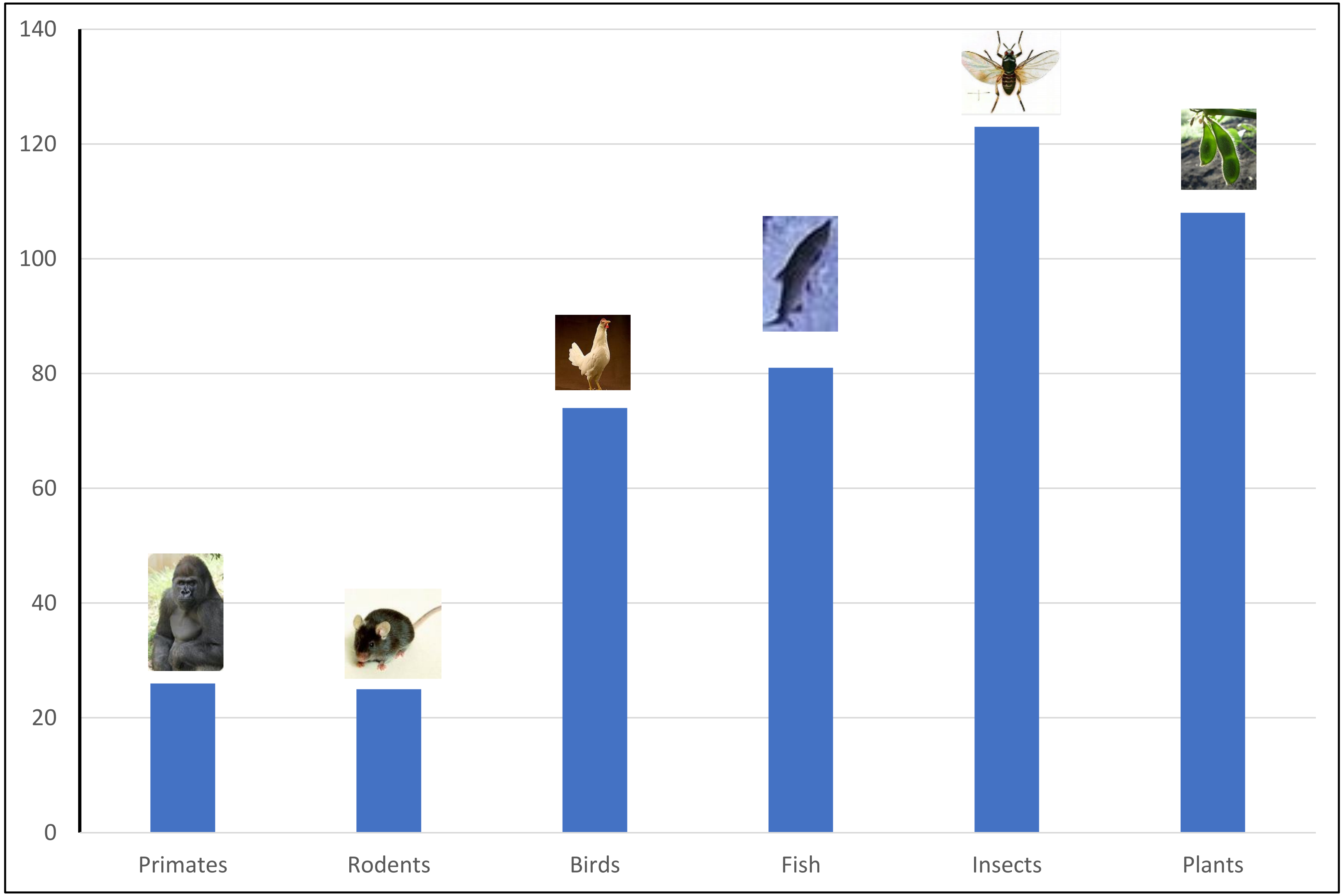


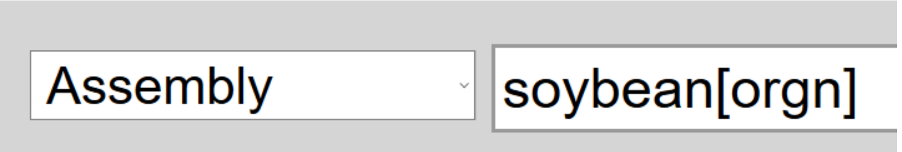
108 plant genomes have been annotated at NCBI



https://www.ncbi.nlm.nih.gov/taxonomy
https://www.ncbi.nlm.nih.gov/tools/treeviewer/

Annotation of various organism groups



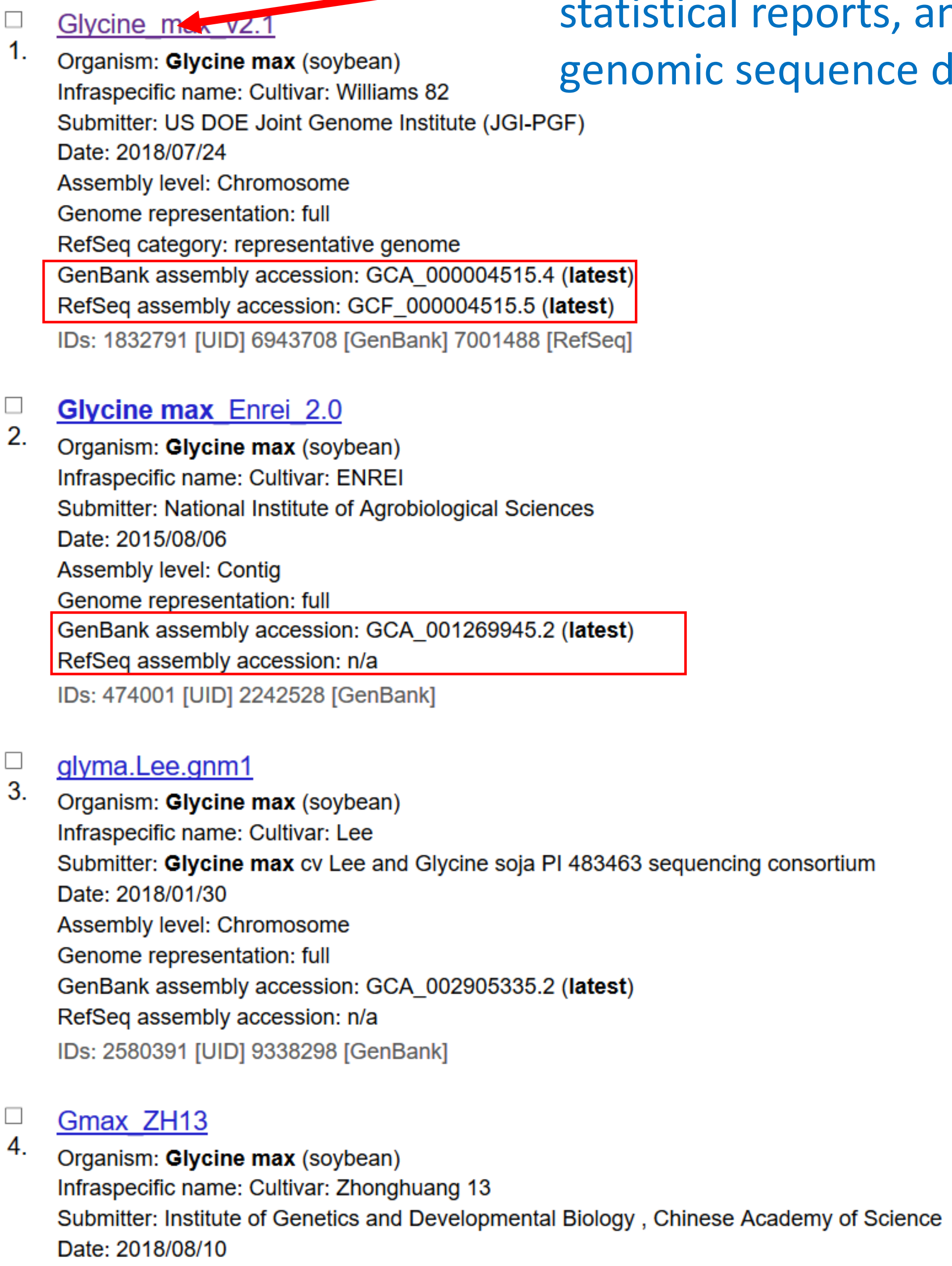


Assembly

soybean[orgn]

Best representative assembly chosen for annotation

Access assembly meta-data, statistical reports, and links to genomic sequence data



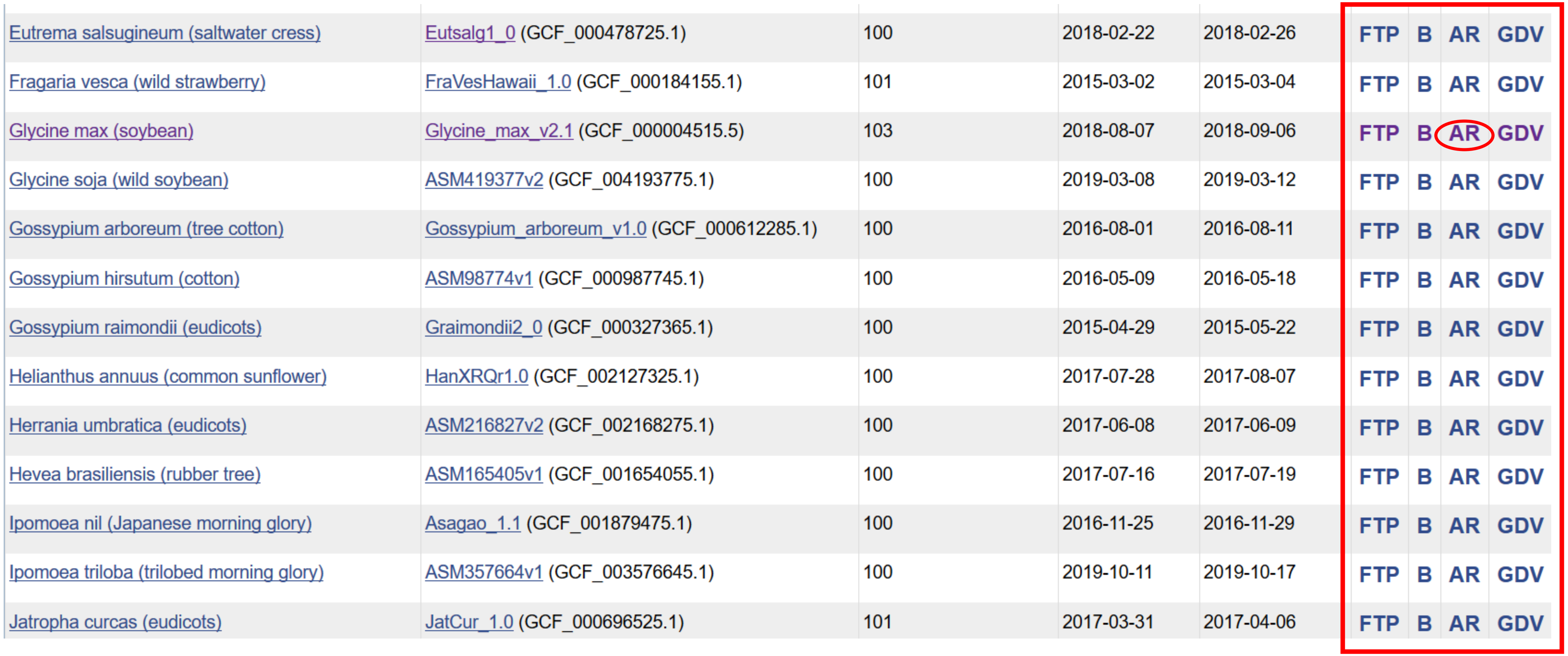
□ 1. Glycine_max_v2.1
Organism: **Glycine max** (soybean)
Infraspecific name: Cultivar: Williams 82
Submitter: US DOE Joint Genome Institute (JGI-PGF)
Date: 2018/07/24
Assembly level: Chromosome
Genome representation: full
RefSeq category: representative genome
GenBank assembly accession: GCA_000004515.4 (**latest**)
RefSeq assembly accession: GCF_000004515.5 (**latest**)
IDs: 1832791 [UID] 6943708 [GenBank] 7001488 [RefSeq]

□ 2. **Glycine max**_Enrei_2.0
Organism: **Glycine max** (soybean)
Infraspecific name: Cultivar: ENREI
Submitter: National Institute of Agrobiological Sciences
Date: 2015/08/06
Assembly level: Contig
Genome representation: full
GenBank assembly accession: GCA_001269945.2 (**latest**)
RefSeq assembly accession: n/a
IDs: 474001 [UID] 2242528 [GenBank]

□ 3. glyma.Lee.gnm1
Organism: **Glycine max** (soybean)
Infraspecific name: Cultivar: Lee
Submitter: **Glycine max** cv Lee and Glycine soja PI 483463 sequencing consortium
Date: 2018/01/30
Assembly level: Chromosome
Genome representation: full
GenBank assembly accession: GCA_002905335.2 (**latest**)
RefSeq assembly accession: n/a
IDs: 2580391 [UID] 9338298 [GenBank]

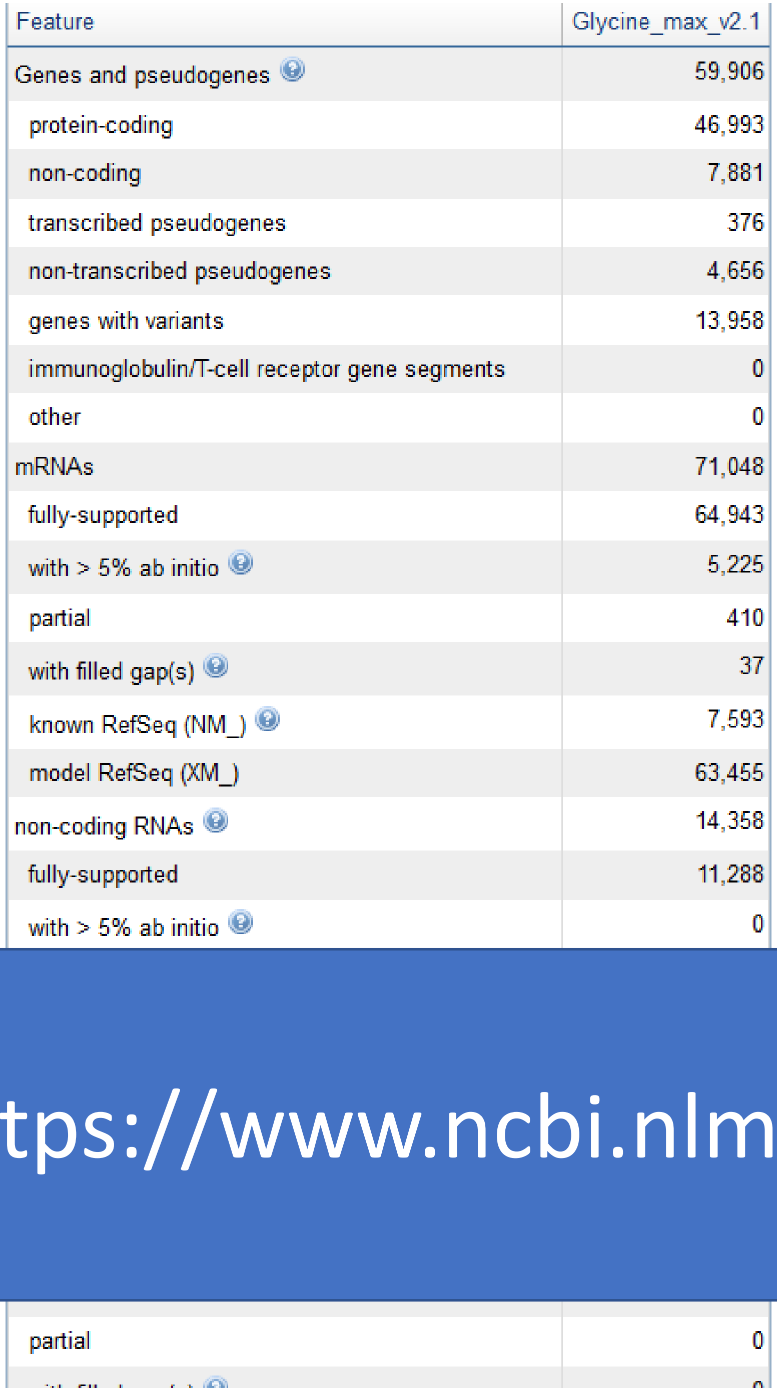
□ 4. Gmax_ZH13
Organism: **Glycine max** (soybean)
Infraspecific name: Cultivar: Zhonghuang 13
Submitter: Institute of Genetics and Developmental Biology , Chinese Academy of Science
Date: 2018/08/10

Annotation Home Page

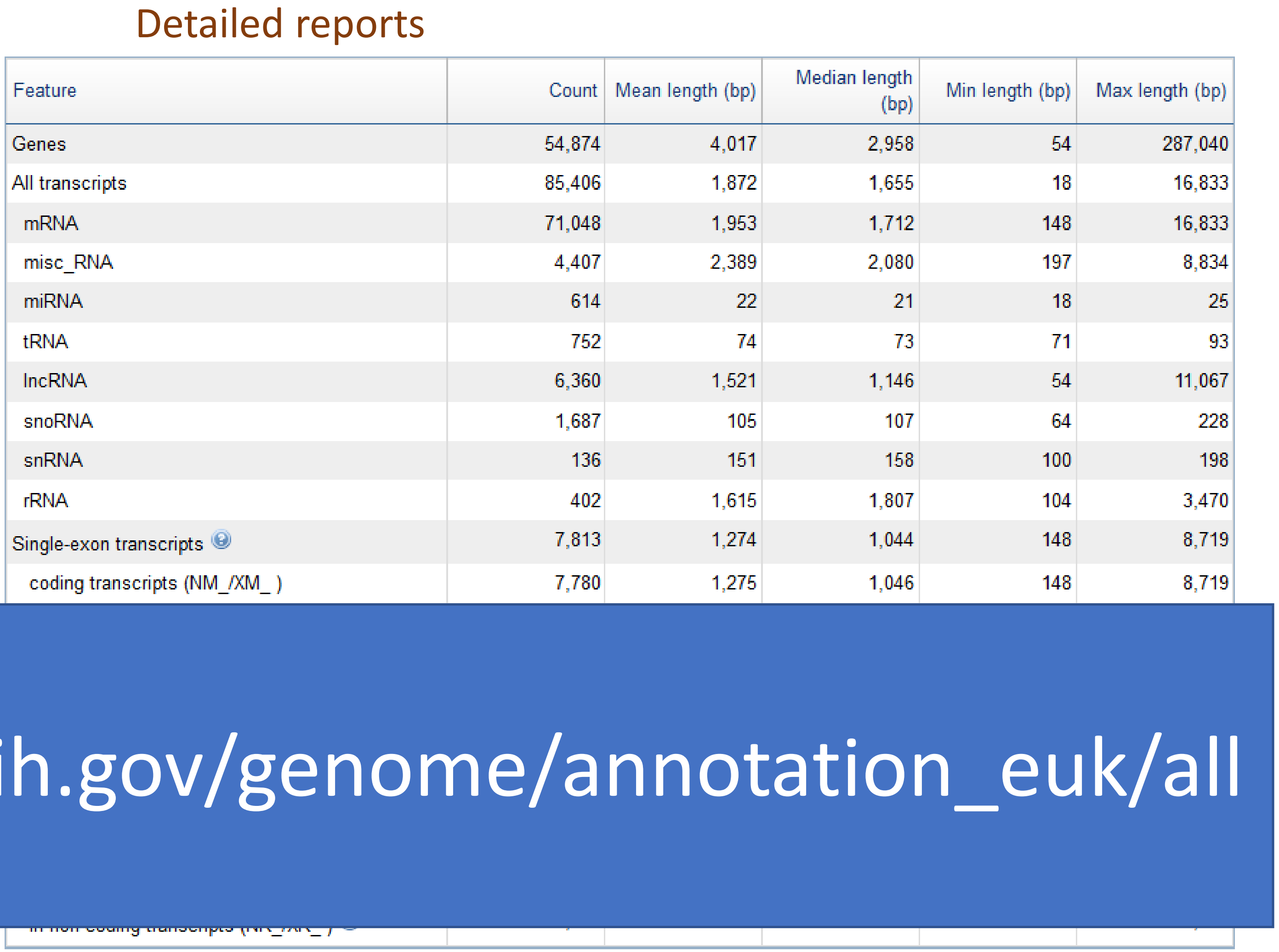


Eutrema salsugineum (saltwater cress)	Eutsalg1_0 (GCF_000478725.1)	100	2018-02-22	2018-02-26	FTP	B	AR	GDV
Fragaria vesca (wild strawberry)	FraVesHawaii_1.0 (GCF_000184155.1)	101	2015-03-02	2015-03-04	FTP	B	AR	GDV
Glycine max (soybean)	Glycine_max_v2.1 (GCF_000004515.5)	103	2018-08-07	2018-09-06	FTP	B	AR	GDV
Glycine soja (wild soybean)	ASM419377v2 (GCF_004193775.1)	100	2019-03-08	2019-03-12	FTP	B	AR	GDV
Gossypium arboreum (tree cotton)	Gossypium_arboreum_v1.0 (GCF_000612285.1)	100	2016-08-01	2016-08-11	FTP	B	AR	GDV
Gossypium hirsutum (cotton)	ASM98774v1 (GCF_000987745.1)	100	2016-05-09	2016-05-18	FTP	B	AR	GDV
Gossypium raimondii (eudicots)	Graimondii2_0 (GCF_000327365.1)	100	2015-04-29	2015-05-22	FTP	B	AR	GDV
Helianthus annuus (common sunflower)	HanXRQr1.0 (GCF_002127325.1)	100	2017-07-28	2017-08-07	FTP	B	AR	GDV
Herrania umbratica (eudicots)	ASM216827v2 (GCF_002168275.1)	100	2017-06-08	2017-06-09	FTP	B	AR	GDV
Hevea brasiliensis (rubber tree)	ASM165405v1 (GCF_001654055.1)	100	2017-07-16	2017-07-19	FTP	B	AR	GDV
Ipomoea nil (Japanese morning glory)	Asagao_1.1 (GCF_001879475.1)	100	2016-11-25	2016-11-29	FTP	B	AR	GDV
Ipomoea triloba (trilobed morning glory)	ASM357664v1 (GCF_003576645.1)	100	2019-10-11	2019-10-17	FTP	B	AR	GDV
Jatropha curcas (eudicots)	JatCur_1.0 (GCF_000696525.1)	101	2017-03-31	2017-04-06	FTP	B	AR	GDV

Annotation Report



Feature	Glycine_max_v2.1
Genes and pseudogenes	59,906
protein-coding	46,993
non-coding	7,881
transcribed pseudogenes	376
non-transcribed pseudogenes	4,656
genes with variants	13,958
immunoglobulin/T-cell receptor gene segments	0
other	0
mRNAs	71,048
fully-supported	64,943
with > 5% ab initio	5,225
partial	410
with filled gap(s)	37
known RefSeq (NM_)	7,593
model RefSeq (XM_)	63,455
non-coding RNAs	14,358
fully-supported	11,288
with > 5% ab initio	0
partial	0
with filled gap(s)	0



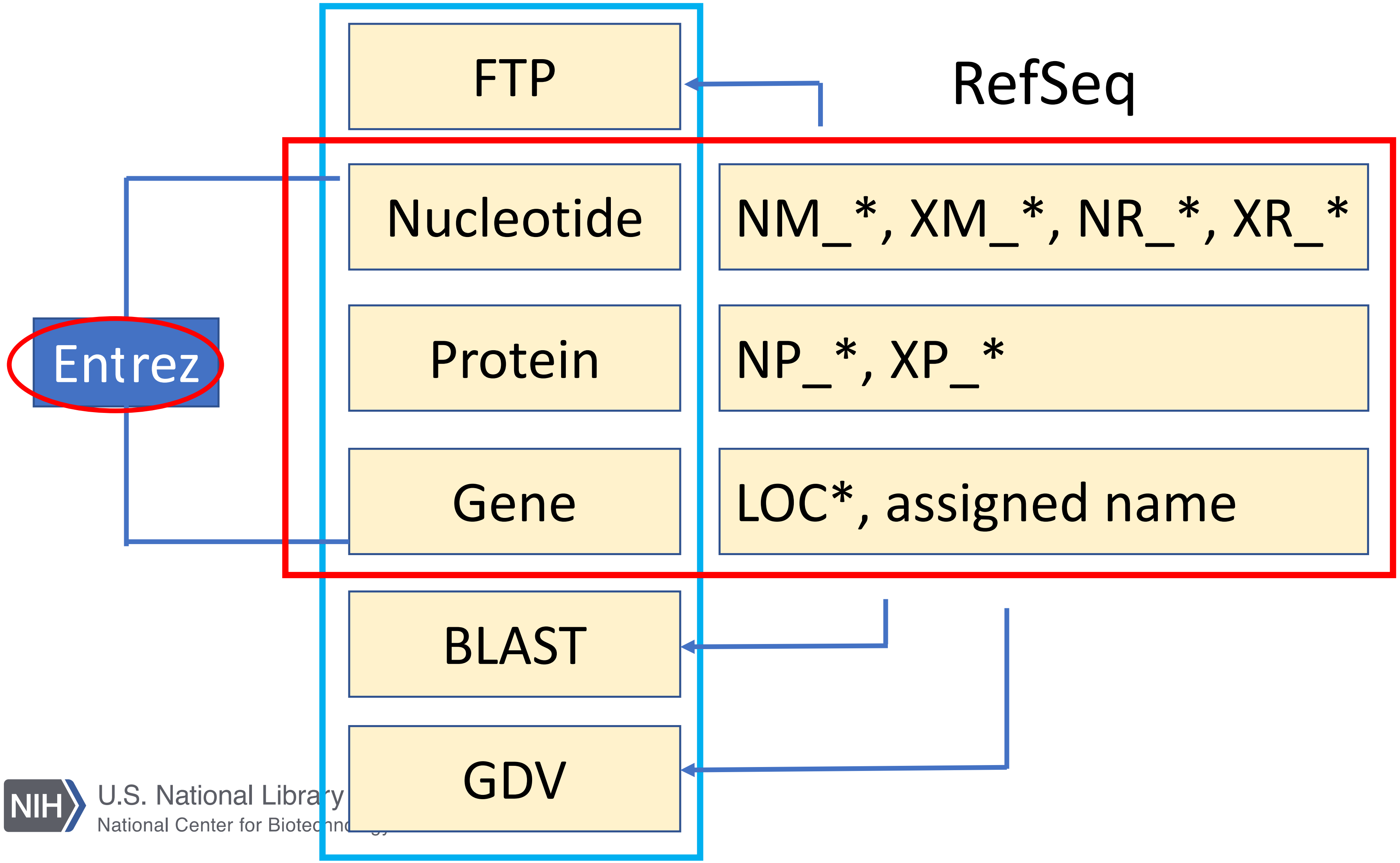
Detailed reports

Feature	Count	Mean length (bp)	Median length (bp)	Min length (bp)	Max length (bp)
Genes	54,874	4,017	2,958	54	287,040
All transcripts	85,406	1,872	1,655	18	16,833
mRNA	71,048	1,953	1,712	148	16,833
misc_RNA	4,407	2,389	2,080	197	8,834
miRNA	614	22	21	18	25
tRNA	752	74	73	71	93
lncRNA	6,360	1,521	1,146	54	11,067
snoRNA	1,687	105	107	64	228
snRNA	136	151	158	100	198
rRNA	402	1,615	1,807	104	3,470
Single-exon transcripts	7,813	1,274	1,044	148	8,719
coding transcripts (NM_/XM_)	7,780	1,275	1,046	148	8,719
[illegible] non-coding transcripts (NR_/XR_)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all

Annotation Output

ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/



RefSeq

Entrez

FTP

Nucleotide — NM_*, XM_*, NR_*, XR_*

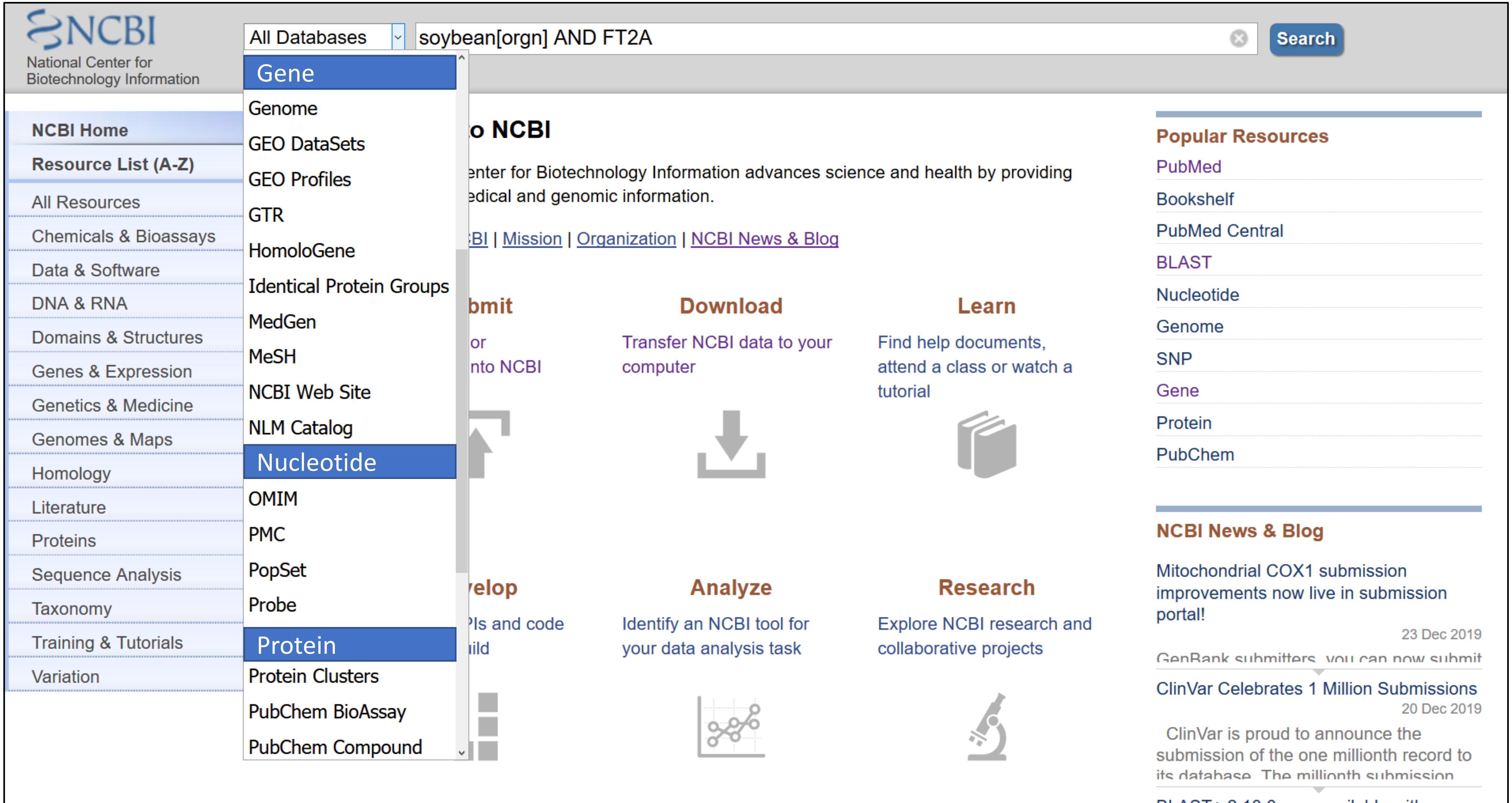
Protein — NP_*, XP_*

Gene — LOC*, assigned name

BLAST

GDV

Data Access-Entrez search



o NCBI

enter for Biotechnology Information advances science and health by providing edical and genomic information.

BI | Mission | Organization | NCBI News & Blog

or nto NCBI

Transfer NCBI data to your computer

Find help documents, attend a class or watch a tutorial

Is and code ild

Identify an NCBI tool for your data analysis task

Explore NCBI research and collaborative projects

NCBI News & Blog

Mitochondrial COX1 submission improvements now live in submission portal!

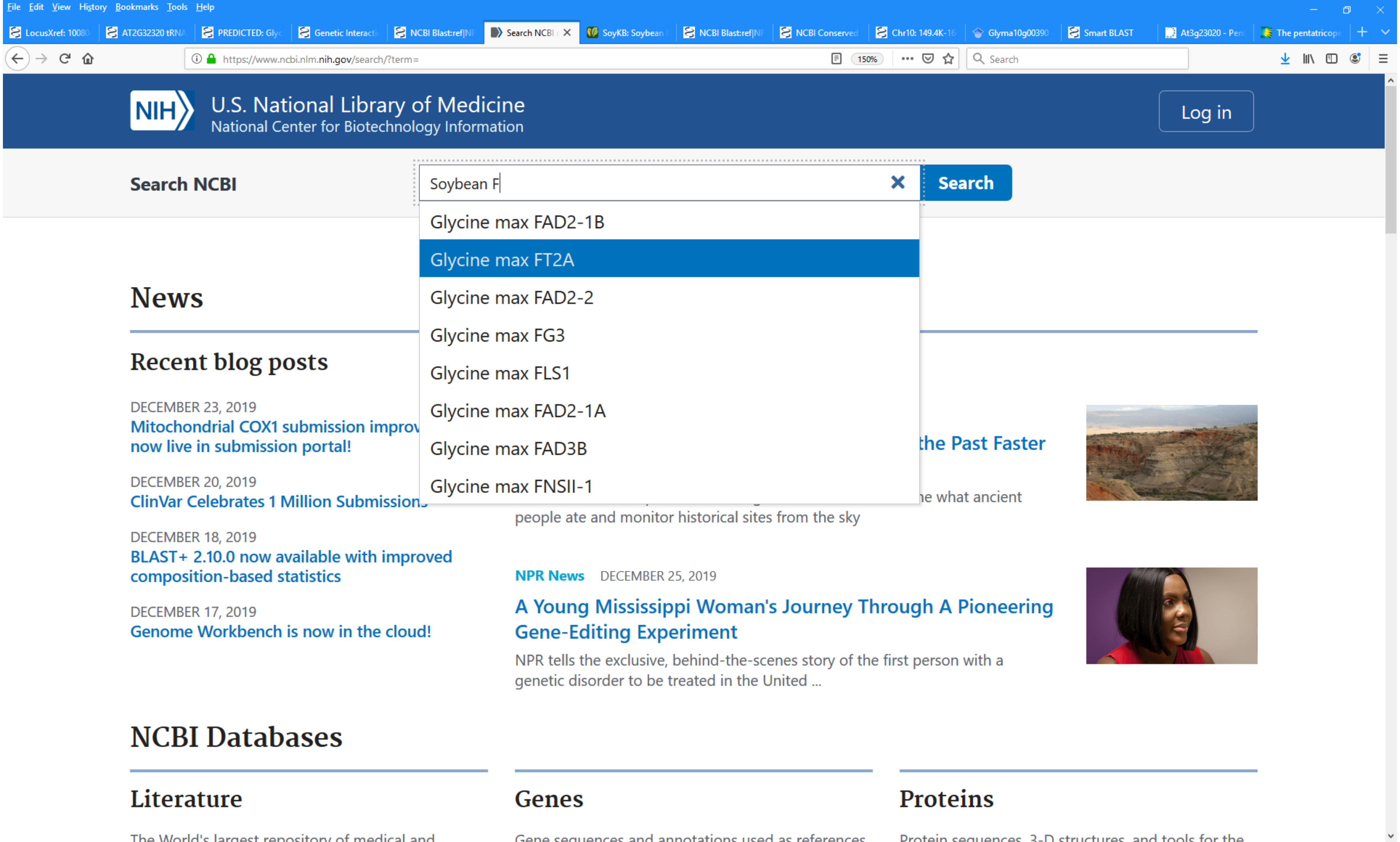
23 Dec 2019

GenBank submitters, you can now submit

ClinVar Celebrates 1 Million Submissions

20 Dec 2019

ClinVar is proud to announce the submission of the one millionth record to its database. The millionth submission



U.S. National Library of Medicine
National Center for Biotechnology Information

Search NCBI

Soybean F

Search

- Glycine max FAD2-1B
- Glycine max FT2A
- Glycine max FAD2-2
- Glycine max FG3
- Glycine max FLS1
- Glycine max FAD2-1A
- Glycine max FAD3B
- Glycine max FNSII-1

News

Recent blog posts

DECEMBER 23, 2019
Mitochondrial COX1 submission improv
now live in submission portal!

DECEMBER 20, 2019
ClinVar Celebrates 1 Million Submissions

DECEMBER 18, 2019
BLAST+ 2.10.0 now available with improved composition-based statistics

DECEMBER 17, 2019
Genome Workbench is now in the cloud!

the Past Faster

he what ancient people ate and monitor historical sites from the sky

NPR News DECEMBER 25, 2019
A Young Mississippi Woman's Journey Through A Pioneering Gene-Editing Experiment

NPR tells the exclusive, behind-the-scenes story of the first person with a genetic disorder to be treated in the United ...

NCBI Databases

Literature

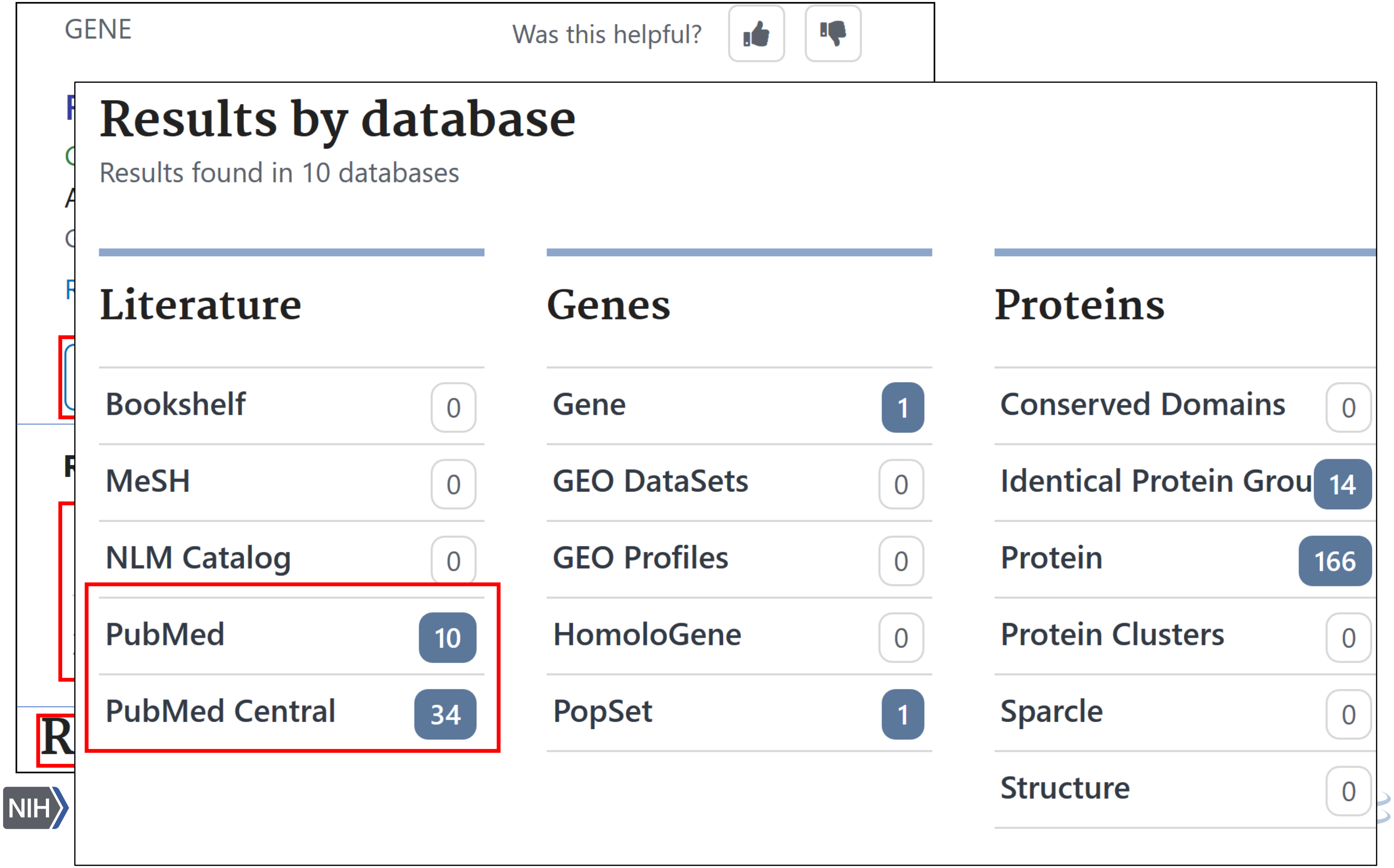
The World's largest repository of medical and

Genes

Gene sequences and annotations used as references

Proteins

Protein sequences, 3-D structures, and tools for the



GENE

Was this helpful?

Results by database

Results found in 10 databases

Literature

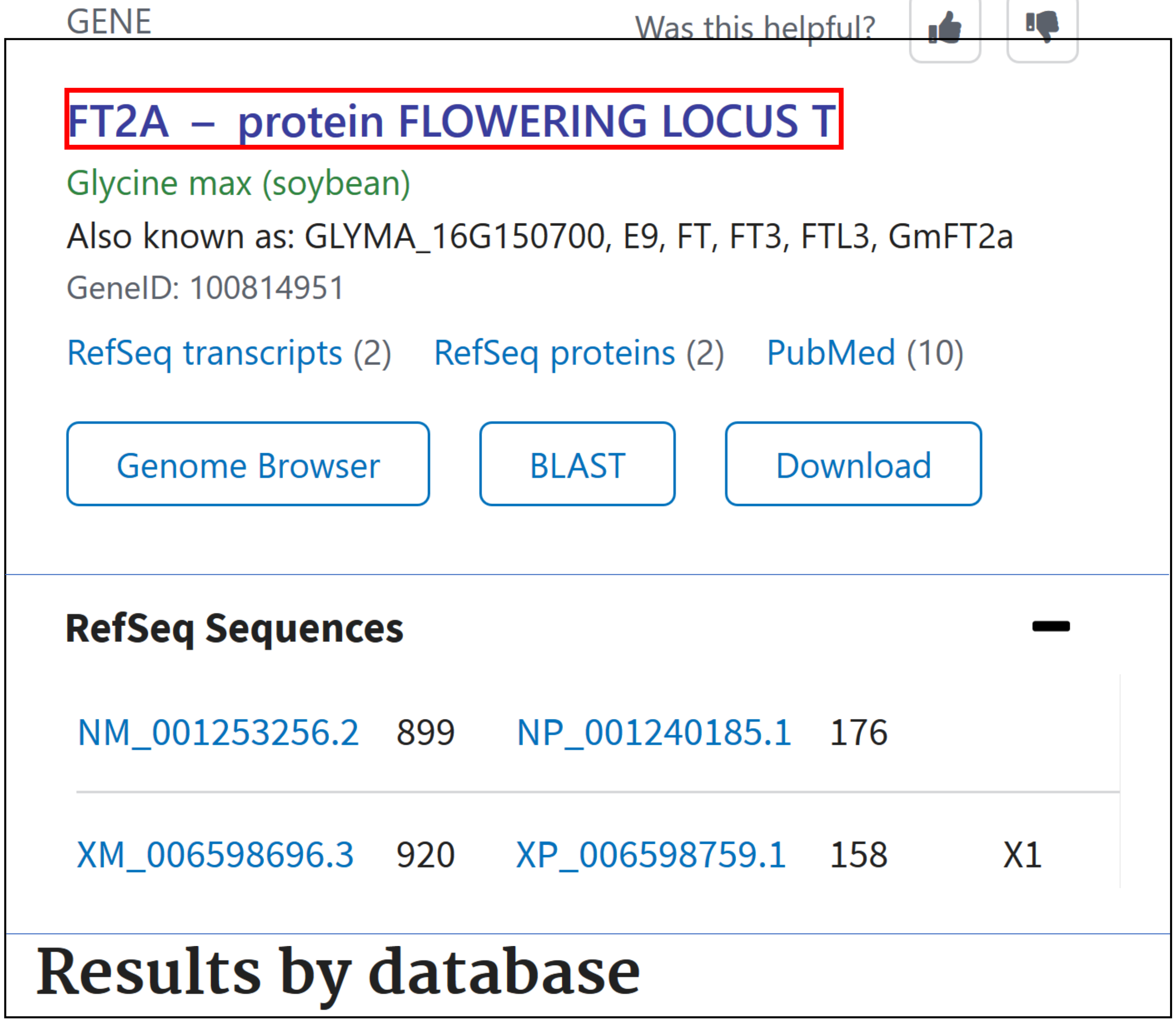
Bookshelf	0
MeSH	0
NLM Catalog	0
PubMed	10
PubMed Central	34

Genes

Gene	1
GEO DataSets	0
GEO Profiles	0
HomoloGene	0
PopSet	1

Proteins

Conserved Domains	0
Identical Protein Grou	14
Protein	166
Protein Clusters	0
Sparcle	0
Structure	0



FT2A – protein FLOWERING LOCUS T

Glycine max (soybean)

Also known as: GLYMA_16G150700, E9, FT, FT3, FTL3, GmFT2a

GeneID: 100814951

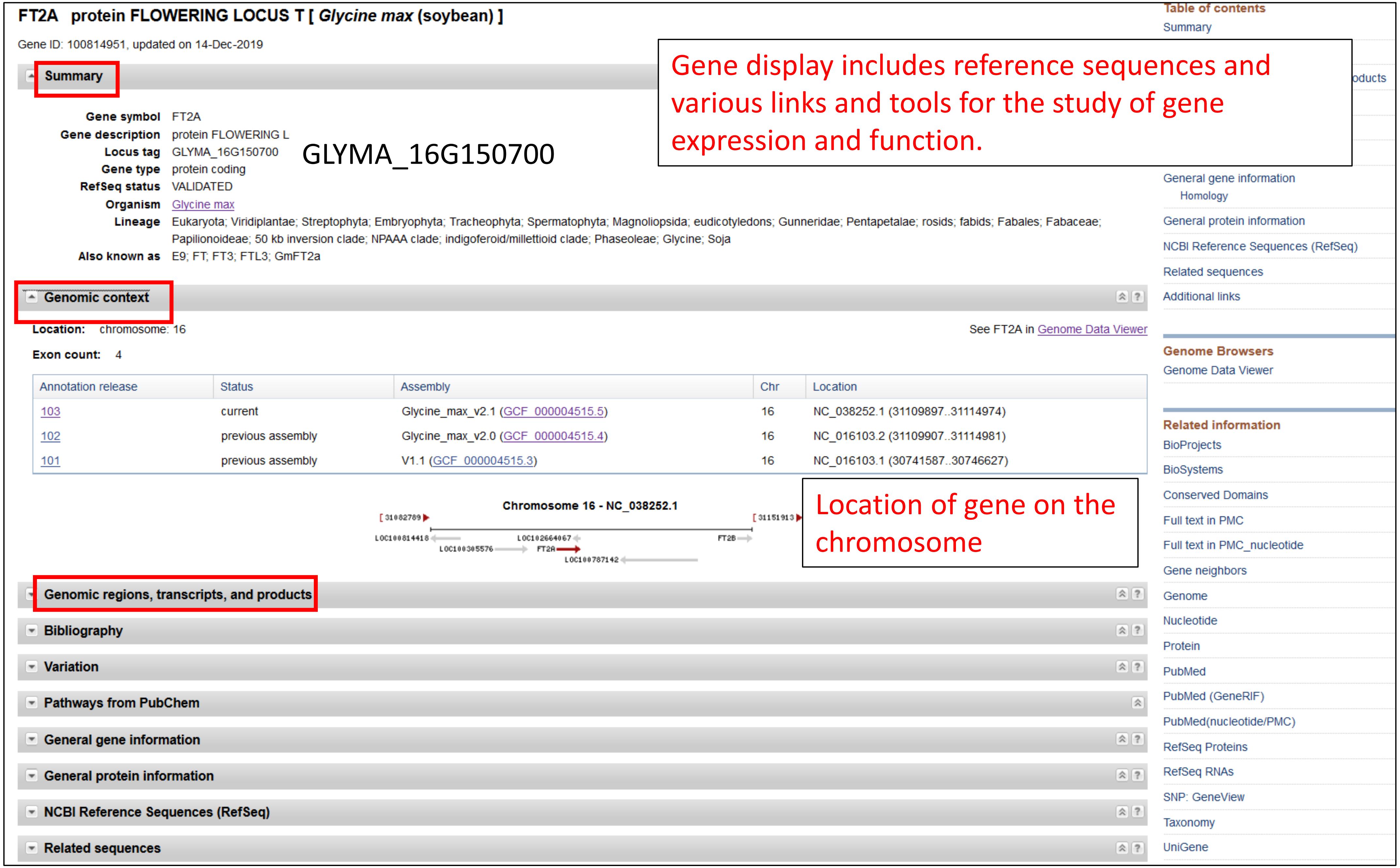
RefSeq transcripts (2) RefSeq proteins (2) PubMed (10)

RefSeq Sequences

NM_001253256.2	899	NP_001240185.1	176	
XM_006598696.3	920	XP_006598759.1	158	X1

Results by database

Data access – Gene and RefSeq



FT2A protein FLOWERING LOCUS T [*Glycine max* (soybean)]

Gene ID: 100814951, updated on 14-Dec-2019

Summary

Gene symbol	FT2A
Gene description	protein FLOWERING L
Locus tag	GLYMA_16G150700
Gene type	protein coding
RefSeq status	VALIDATED
Organism	Glycine max
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade; NPAAA clade; indigoferoid/millettioid clade; Phaseoleae; Glycine; Soja
Also known as	E9; FT; FT3; FTL3; GmFT2a

GLYMA_16G150700

Gene display includes reference sequences and various links and tools for the study of gene expression and function.

Genomic context

Location: chromosome: 16

See FT2A in Genome Data Viewer

Exon count: 4

Annotation release	Status	Assembly	Chr	Location
103	current	Glycine_max_v2.1 (GCF_000004515.5)	16	NC_038252.1 (31109897..31114974)
102	previous assembly	Glycine_max_v2.0 (GCF_000004515.4)	16	NC_016103.2 (31109907..31114981)
101	previous assembly	V1.1 (GCF_000004515.3)	16	NC_016103.1 (30741587..30746627)

Location of gene on the chromosome

Genomic regions, transcripts, and products

Bibliography

Variation

Pathways from PubChem

General gene information

General protein information

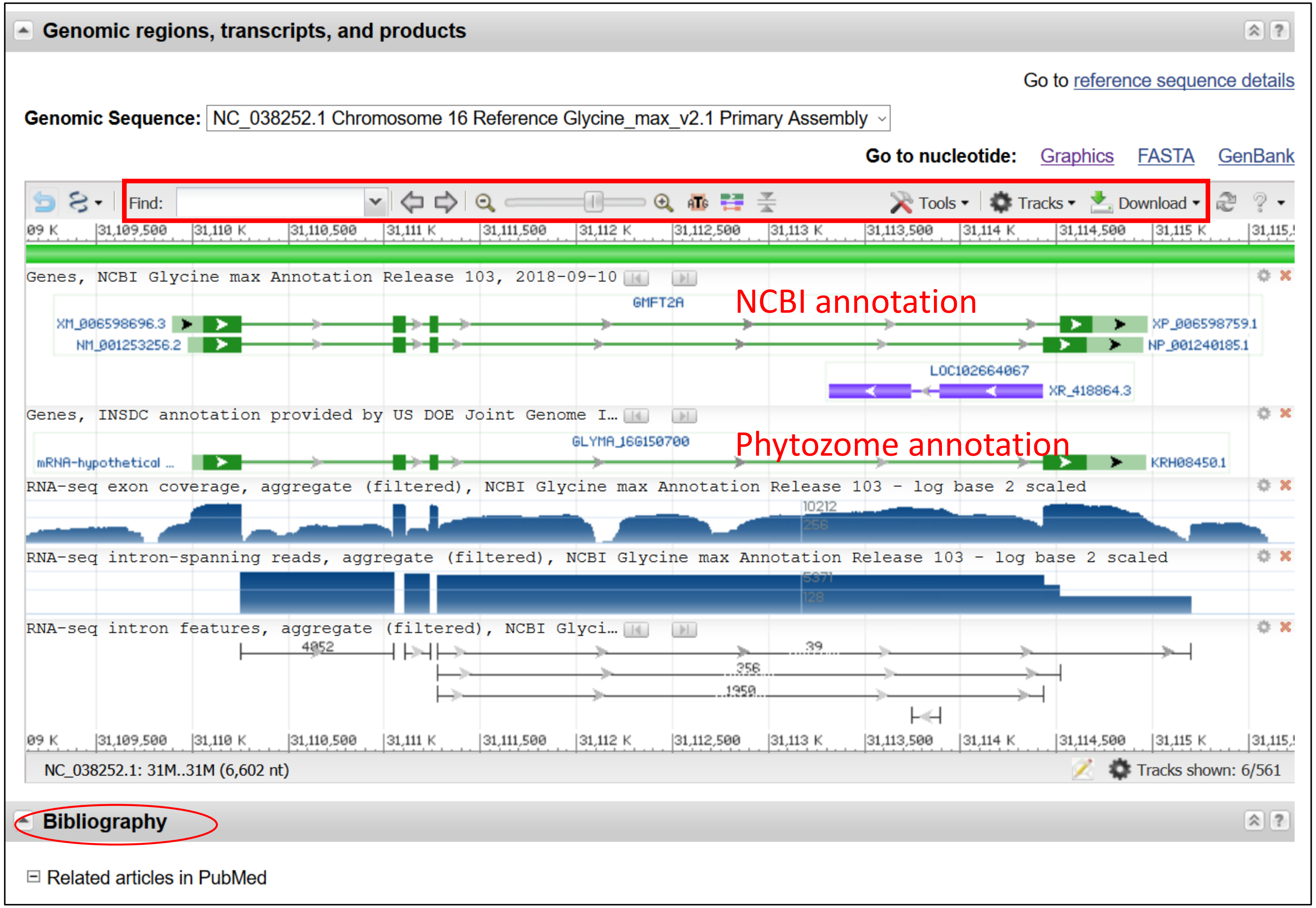
NCBI Reference Sequences (RefSeq)

Related sequences

Table of contents
Summary
oducts
General gene information
Homology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links

Genome Browsers
Genome Data Viewer

Related information
BioProjects
BioSystems
Conserved Domains
Full text in PMC
Full text in PMC_nucleotide
Gene neighbors
Genome
Nucleotide
Protein
PubMed
PubMed (GeneRIF)
PubMed(nucleotide/PMC)
RefSeq Proteins
RefSeq RNAs
SNP: GeneView
Taxonomy
UniGene



Genes, NCBI Glycine max Annotation Release 103, 2018-09-10
GMFT2A
XM_006598696.3
NM_001253256.2
XP_006598759.1
NP_001240185.1
LOC102664067
XR_418864.3
Genes, INSDC annotation provided by US DOE Joint Genome I...
GLYMA_16G150700
mRNA-hypothetical ...
KRH08450.1
RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
10212
256
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
5371
128
RNA-seq intron features, aggregate (filtered), NCBI Glyci...
4052
39
356
1950
09 K
31,109,500
31,110 K
31,110,500
31,111 K
31,111,500
31,112 K
31,112,500
31,113 K
31,113,500
31,114 K
31,114,500
31,115 K
31,115,
Find:
Tools
Tracks
Download
Graphics
FASTA
GenBank

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_038252.1 Chromosome 16 Reference Glycine_max_v2.1 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

NCBI annotation

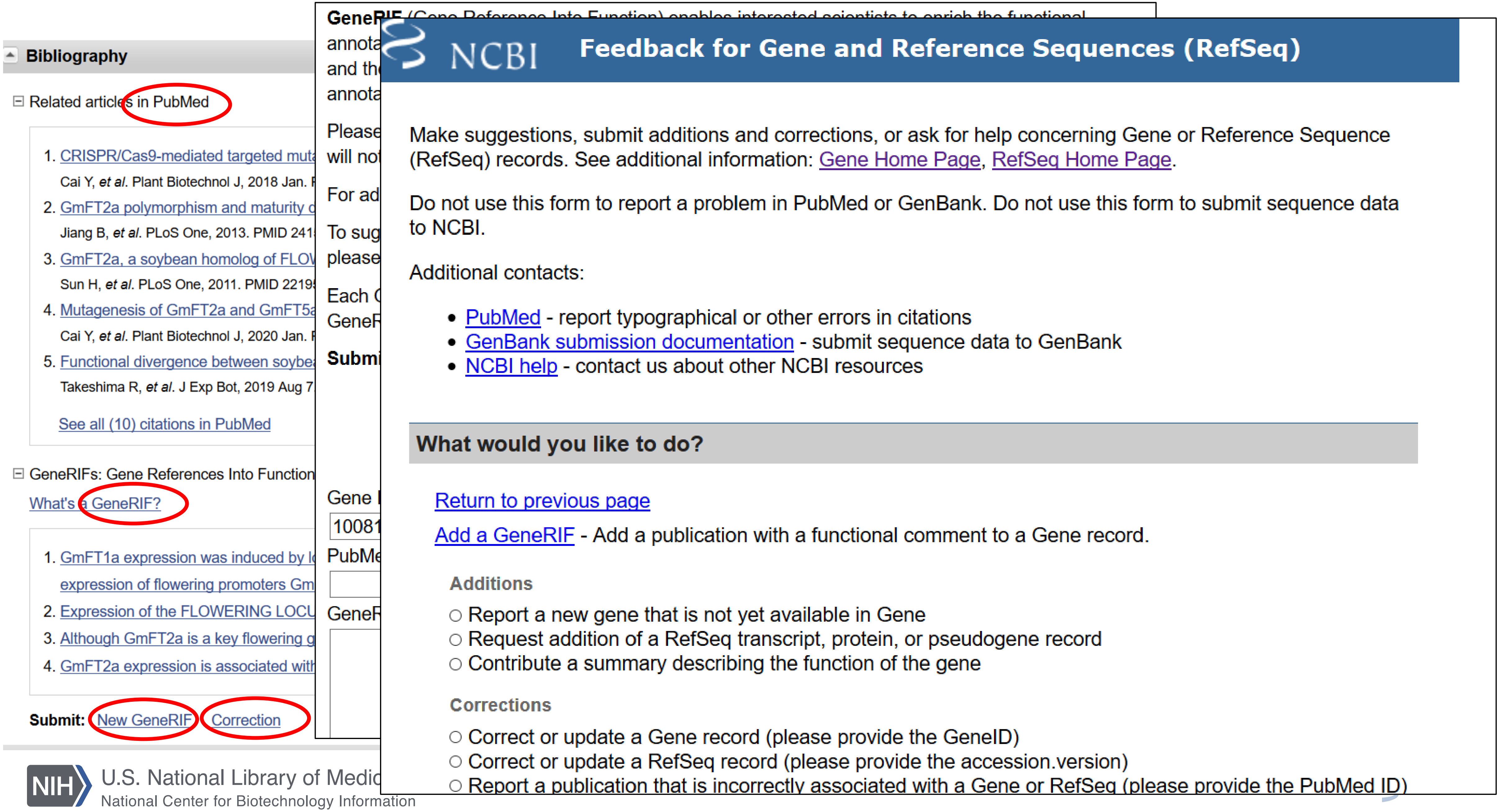
Phytozome annotation

NC_038252.1: 31M..31M (6,602 nt)

Tracks shown: 6/561

Bibliography

⊟ Related articles in PubMed



Bibliography

⊟ Related articles in PubMed

1. CRISPR/Cas9-mediated targeted muta
   Cai Y, *et al*. Plant Biotechnol J, 2018 Jan. P
2. GmFT2a polymorphism and maturity c
   Jiang B, *et al*. PLoS One, 2013. PMID 241
3. GmFT2a, a soybean homolog of FLOW
   Sun H, *et al*. PLoS One, 2011. PMID 2219
4. Mutagenesis of GmFT2a and GmFT5a
   Cai Y, *et al*. Plant Biotechnol J, 2020 Jan. P
5. Functional divergence between soybe
   Takeshima R, *et al*. J Exp Bot, 2019 Aug 7

See all (10) citations in PubMed

⊟ GeneRIFs: Gene References Into Function

What's a GeneRIF?

1. GmFT1a expression was induced by l
   expression of flowering promoters Gm
2. Expression of the FLOWERING LOCU
3. Although GmFT2a is a key flowering g
4. GmFT2a expression is associated with

Submit: New GeneRIF Correction

U.S. National Library of Medic
National Center for Biotechnology Information

GeneRIF (Gene Reference Into Function) enables interested scientists to enrich the functional
annota
and th
annota

Please
will not

For ad

To sug
please

Each G
GeneR

Submi

Gene I
10081

PubMe

GeneR

Feedback for Gene and Reference Sequences (RefSeq)

Make suggestions, submit additions and corrections, or ask for help concerning Gene or Reference Sequence (RefSeq) records. See additional information: Gene Home Page, RefSeq Home Page.

Do not use this form to report a problem in PubMed or GenBank. Do not use this form to submit sequence data to NCBI.

Additional contacts:

- PubMed - report typographical or other errors in citations
- GenBank submission documentation - submit sequence data to GenBank
- NCBI help - contact us about other NCBI resources

What would you like to do?

Return to previous page

Add a GeneRIF - Add a publication with a functional comment to a Gene record.

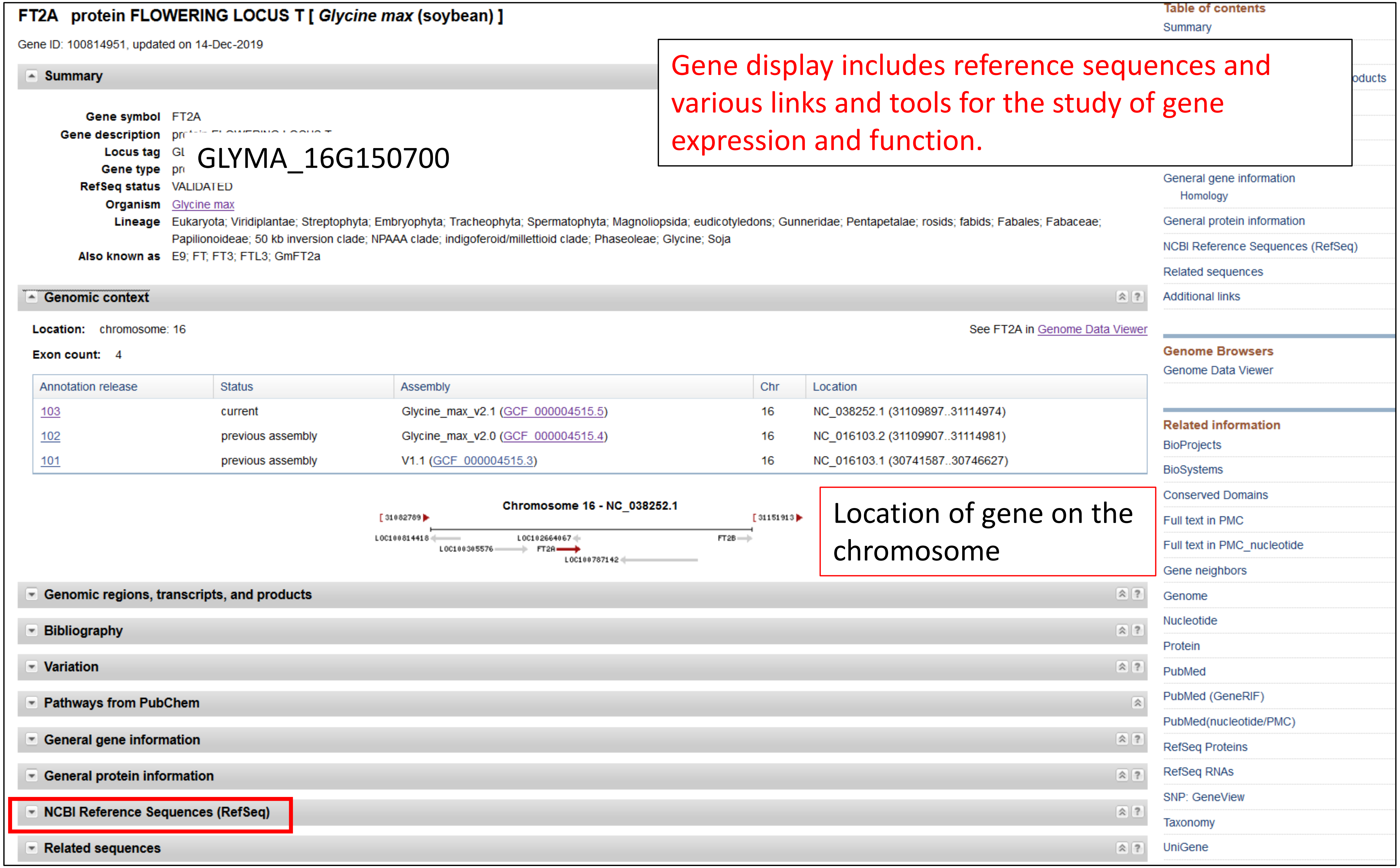
Additions

○ Report a new gene that is not yet available in Gene
○ Request addition of a RefSeq transcript, protein, or pseudogene record
○ Contribute a summary describing the function of the gene

Corrections

○ Correct or update a Gene record (please provide the GeneID)
○ Correct or update a RefSeq record (please provide the accession.version)
○ Report a publication that is incorrectly associated with a Gene or RefSeq (please provide the PubMed ID)

Data access – Gene and RefSeq



FT2A protein FLOWERING LOCUS T [*Glycine max* (soybean)]

Gene ID: 100814951, updated on 14-Dec-2019

Summary

Gene symbol	FT2A
Gene description	pr
Locus tag	Gl
Gene type	pr
RefSeq status	VALIDATED
Organism	Glycine max
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade; NPAAA clade; indigoferoid/millettioid clade; Phaseoleae; Glycine; Soja
Also known as	E9; FT; FT3; FTL3; GmFT2a

GLYMA_16G150700

Genomic context

Location: chromosome: 16

See FT2A in Genome Data Viewer

Exon count: 4

Annotation release	Status	Assembly	Chr	Location
103	current	Glycine_max_v2.1 (GCF_000004515.5)	16	NC_038252.1 (31109897..31114974)
102	previous assembly	Glycine_max_v2.0 (GCF_000004515.4)	16	NC_016103.2 (31109907..31114981)
101	previous assembly	V1.1 (GCF_000004515.3)	16	NC_016103.1 (30741587..30746627)

Genomic regions, transcripts, and products

Bibliography

Variation

Pathways from PubChem

General gene information

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Table of contents

Summary

oducts

General gene information

Homology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

Genome Browsers

Genome Data Viewer

Related information

BioProjects

BioSystems

Conserved Domains

Full text in PMC

Full text in PMC_nucleotide

Gene neighbors

Genome

Nucleotide

Protein

PubMed

PubMed (GeneRIF)

PubMed(nucleotide/PMC)

RefSeq Proteins

RefSeq RNAs

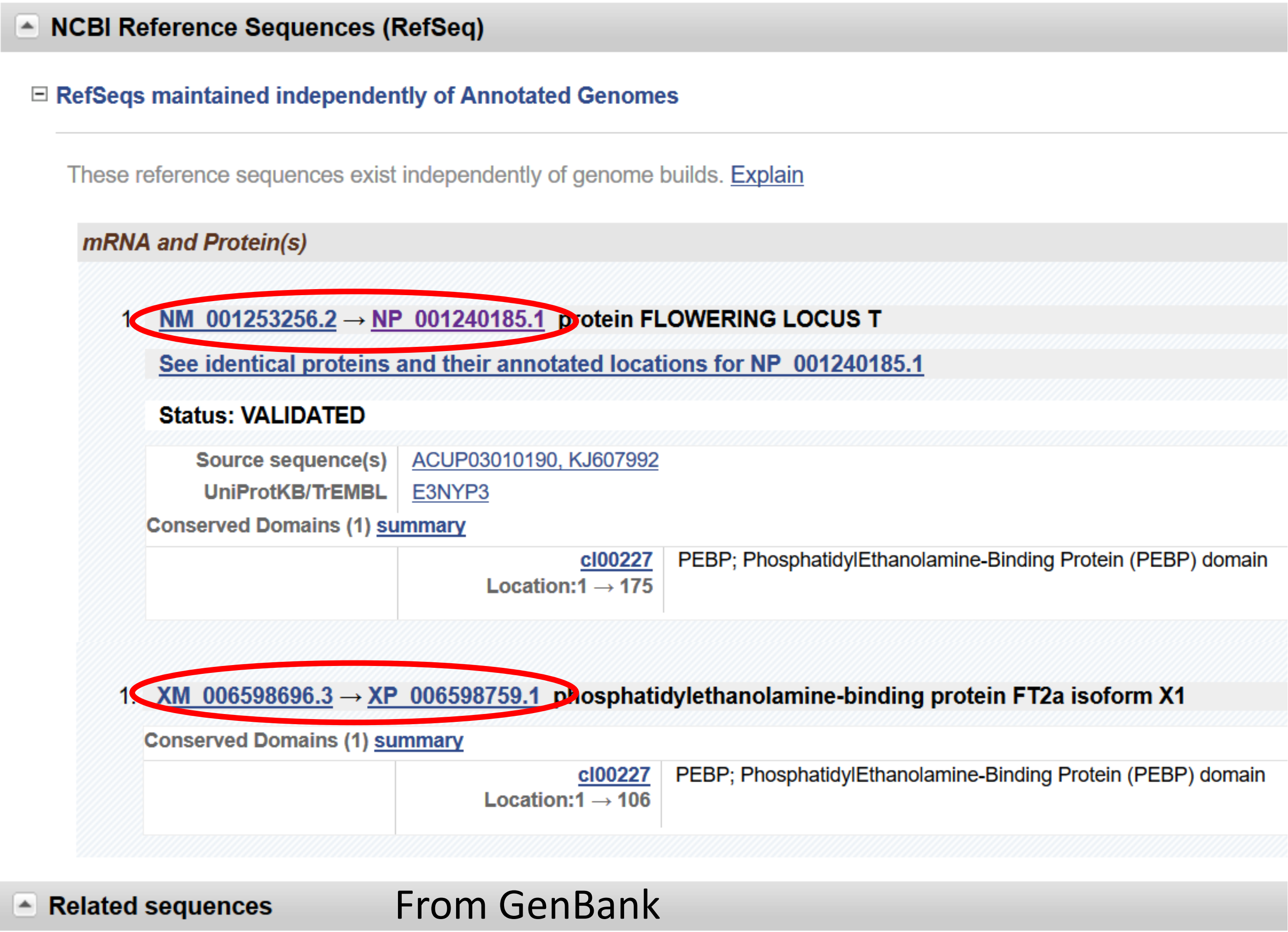
SNP: GeneView

Taxonomy

UniGene

Gene display includes reference sequences and various links and tools for the study of gene expression and function.

Location of gene on the chromosome



NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. Explain

mRNA and Protein(s)

1. NM_001253256.2 → NP_001240185.1 protein FLOWERING LOCUS T

See identical proteins and their annotated locations for NP_001240185.1

Status: VALIDATED

Source sequence(s)	ACUP03010190, KJ607992
UniProtKB/TrEMBL	E3NYP3

Conserved Domains (1) summary

cl00227 Location:1 → 175	PEBP; PhosphatidylEthanolamine-Binding Protein (PEBP) domain

1. XM_006598696.3 → XP_006598759.1 phosphatidylethanolamine-binding protein FT2a isoform X1

Conserved Domains (1) summary

cl00227 Location:1 → 106	PEBP; PhosphatidylEthanolamine-Binding Protein (PEBP) domain

Related sequences

From GenBank

FT2A has two transcript variants

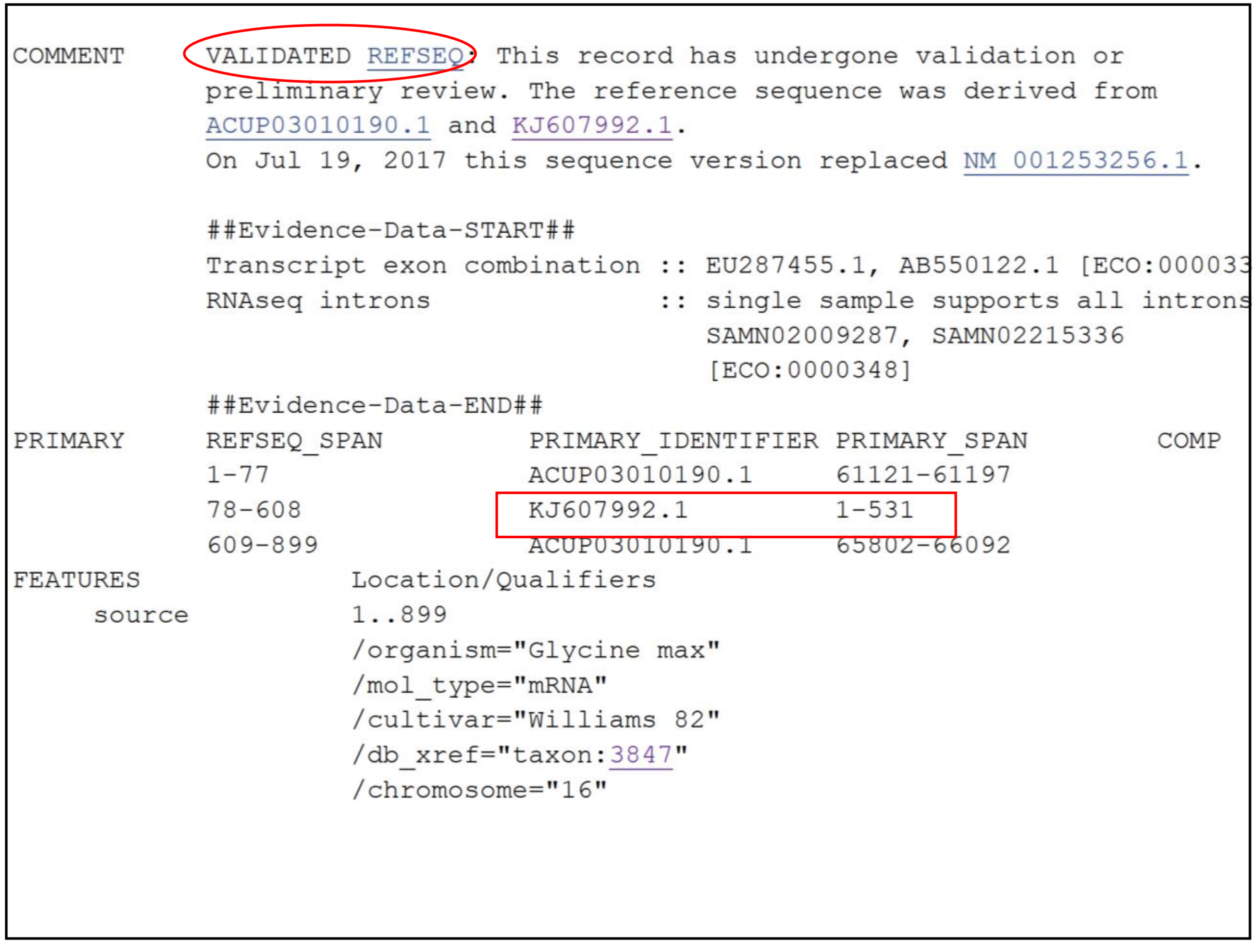
NM_001253256.2
XM_006598696.3

Difference between N* and X* accessions

- N* accessions are used for **known RefSeqs**. This category is supported by manual curation. Records are primarily derived from INSDC cDNAs, EST, and Transcript Shotgun Assembly (TSA) records.

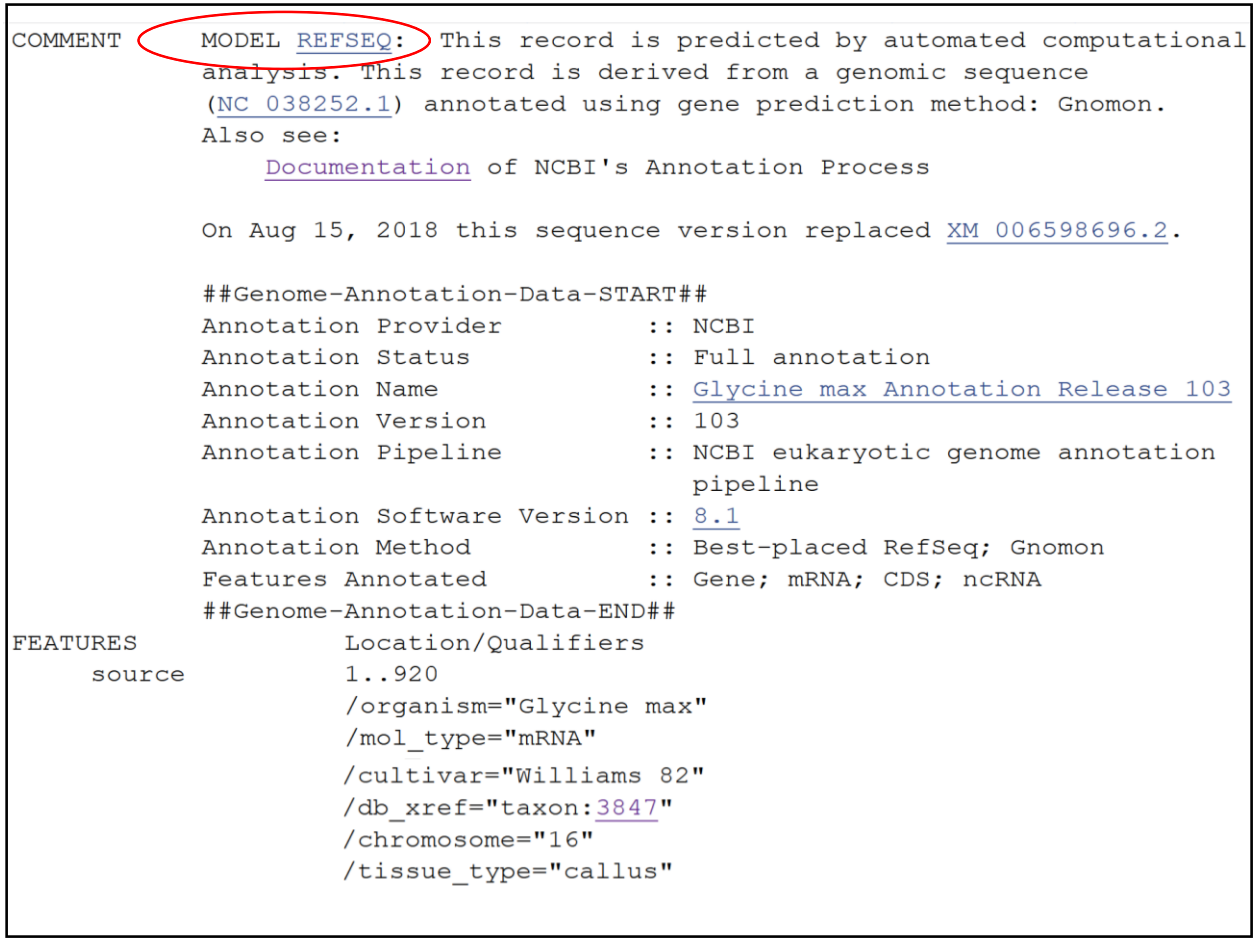
- X* accessions are **predicted models**. This category is computationally predicted based on aligned evidence. Records are primarily derived from genomic sequence. The vast majority of XMs are fully supported by experimental evidence, and for most species **they are on par, quality-wise, with the NMs.**

NM_001253256.2



```
COMMENT     VALIDATED REFSEQ: This record has undergone validation or
            preliminary review. The reference sequence was derived from
            ACUP03010190.1 and KJ607992.1.
            On Jul 19, 2017 this sequence version replaced NM_001253256.1.

            ##Evidence-Data-START##
            Transcript exon combination :: EU287455.1, AB550122.1 [ECO:000033
            RNAseq introns              :: single sample supports all introns
                                           SAMN02009287, SAMN02215336
                                           [ECO:0000348]
            ##Evidence-Data-END##
PRIMARY     REFSEQ_SPAN         PRIMARY_IDENTIFIER PRIMARY_SPAN        COMP
            1-77                ACUP03010190.1     61121-61197
            78-608              KJ607992.1         1-531
            609-899             ACUP03010190.1     65802-66092
FEATURES             Location/Qualifiers
     source          1..899
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /chromosome="16"
```

XM_006598696.3

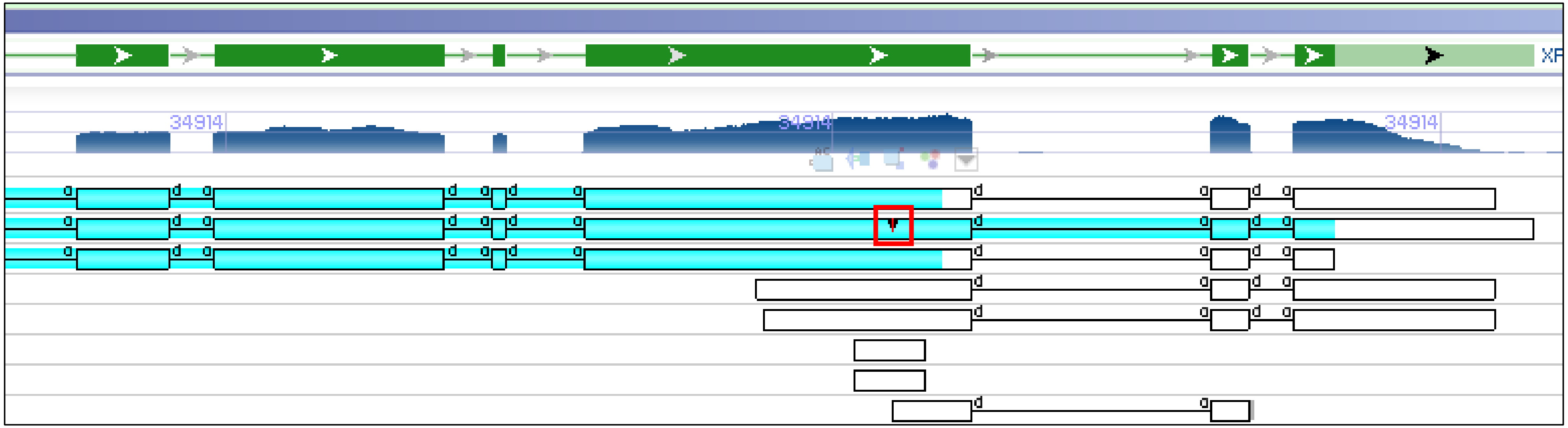


```
COMMENT     MODEL REFSEQ: This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NC_038252.1) annotated using gene prediction method: Gnomon.
            Also see:
                Documentation of NCBI's Annotation Process

            On Aug 15, 2018 this sequence version replaced XM_006598696.2.

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Name             :: Glycine max Annotation Release 103
            Annotation Version          :: 103
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
                                           pipeline
            Annotation Software Version :: 8.1
            Annotation Method           :: Best-placed RefSeq; Gnomon
            Features Annotated          :: Gene; mRNA; CDS; ncRNA
            ##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..920
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /chromosome="16"
                     /tissue_type="callus"
```

Data curation—how do we maintain the quality of our data

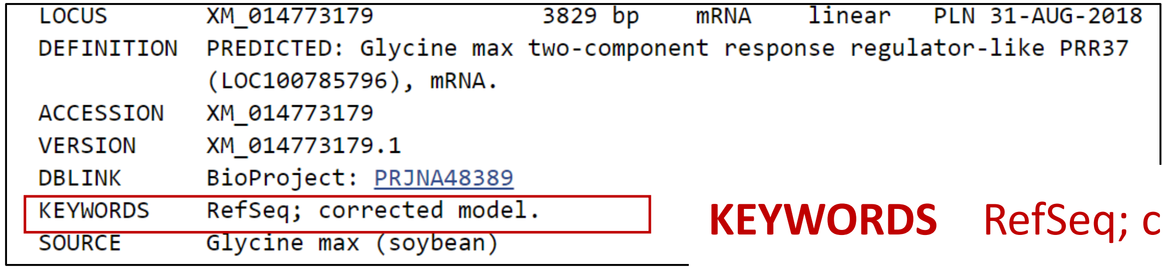
PRR3A gene (GeneID: **100785796)**

Low Quality protein, corrected it based on PMID: 30418611*



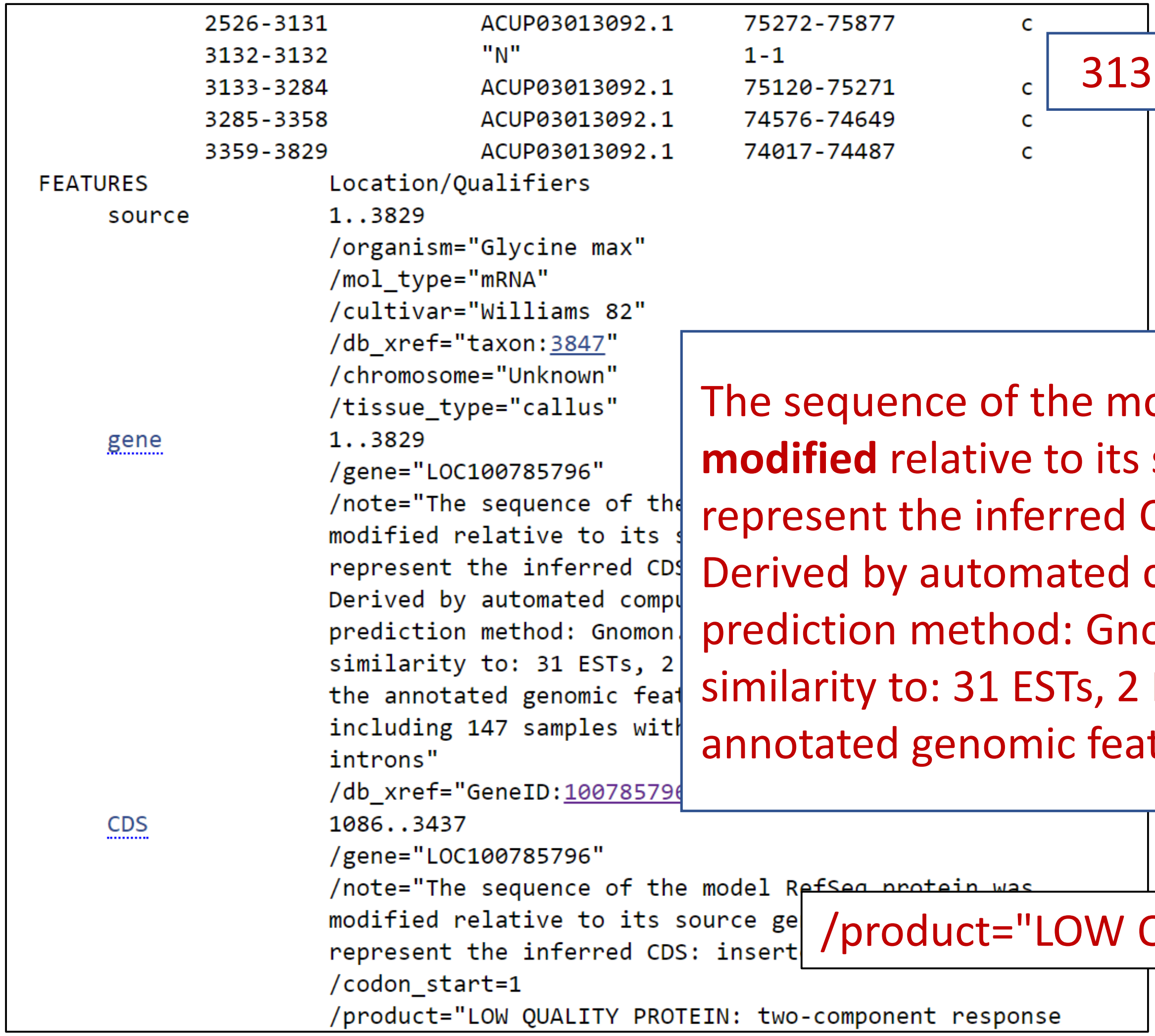
NM_001377264.1
XM_014773179.1

*Li et. al. Plant Cell Physiol. 2019 Feb 1;60(2):407-420. Characterization of Two Growth Period QTLs Reveals Modification of PRR3 Genes During Soybean Domestication.



```
LOCUS       XM_014773179            3829 bp    mRNA    linear   PLN 31-AUG-2018
DEFINITION  PREDICTED: Glycine max two-component response regulator-like PRR37
            (LOC100785796), mRNA.
ACCESSION   XM_014773179
VERSION     XM_014773179.1
DBLINK      BioProject: PRJNA48389
KEYWORDS    RefSeq; corrected model.
SOURCE      Glycine max (soybean)
```

KEYWORDS RefSeq; corrected model



```
            2526-3131           ACUP03013092.1     75272-75877         c
            3132-3132           "N"                1-1
            3133-3284           ACUP03013092.1     75120-75271         c
            3285-3358           ACUP03013092.1     74576-74649         c
            3359-3829           ACUP03013092.1     74017-74487         c
FEATURES             Location/Qualifiers
     source          1..3829
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /chromosome="Unknown"
                     /tissue_type="callus"
     gene            1..3829
                     /gene="LOC100785796"
                     /note="The sequence of the
                     modified relative to its s
                     represent the inferred CDS
                     Derived by automated compu
                     prediction method: Gnomon.
                     similarity to: 31 ESTs, 2
                     the annotated genomic feat
                     including 147 samples with
                     introns"
                     /db_xref="GeneID:10078579
     CDS             1086..3437
                     /gene="LOC100785796"
                     /note="The sequence of the model RefSeq protein was
                     modified relative to its source ge
                     represent the inferred CDS: insert
                     /codon_start=1
                     /product="LOW QUALITY PROTEIN: two-component response
```

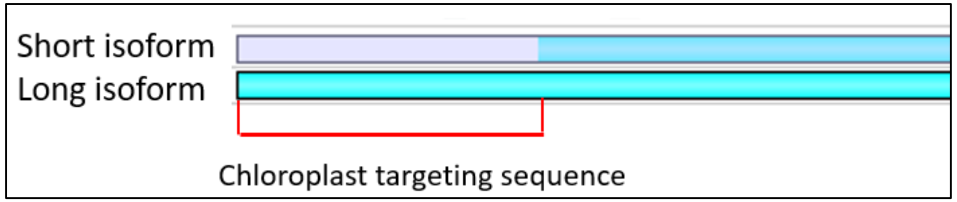
3132-3132 "N" 1-1

The sequence of the model RefSeq **transcript was modified** relative to its source genomic sequence to represent the inferred CDS: **inserted 2 bases in 2 codon**; Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 31 ESTs, 2 Proteins, and 100% coverage of the annotated genomic feature by RNASeq alignments.

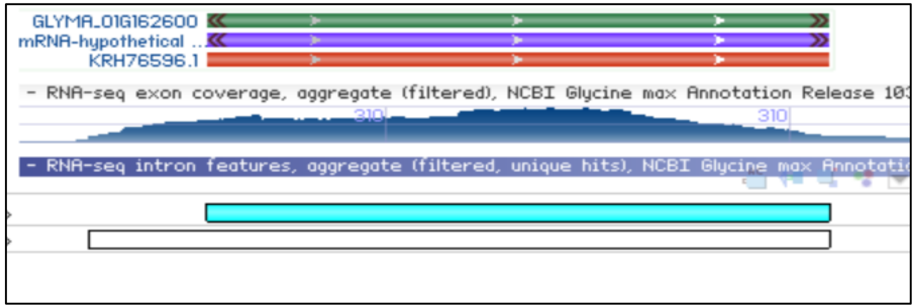
/product="LOW QUALITY PROTEIN: two-component response

Data curation: Add value based on publications

Adding value to data: Updated HPPD gene (GeneID:100101901) to create two isoforms, short and long, based on PMID: 25192697* .



Updating locus_type: Updated LUXa gene from coding to non-coding based on PMID:28878247**

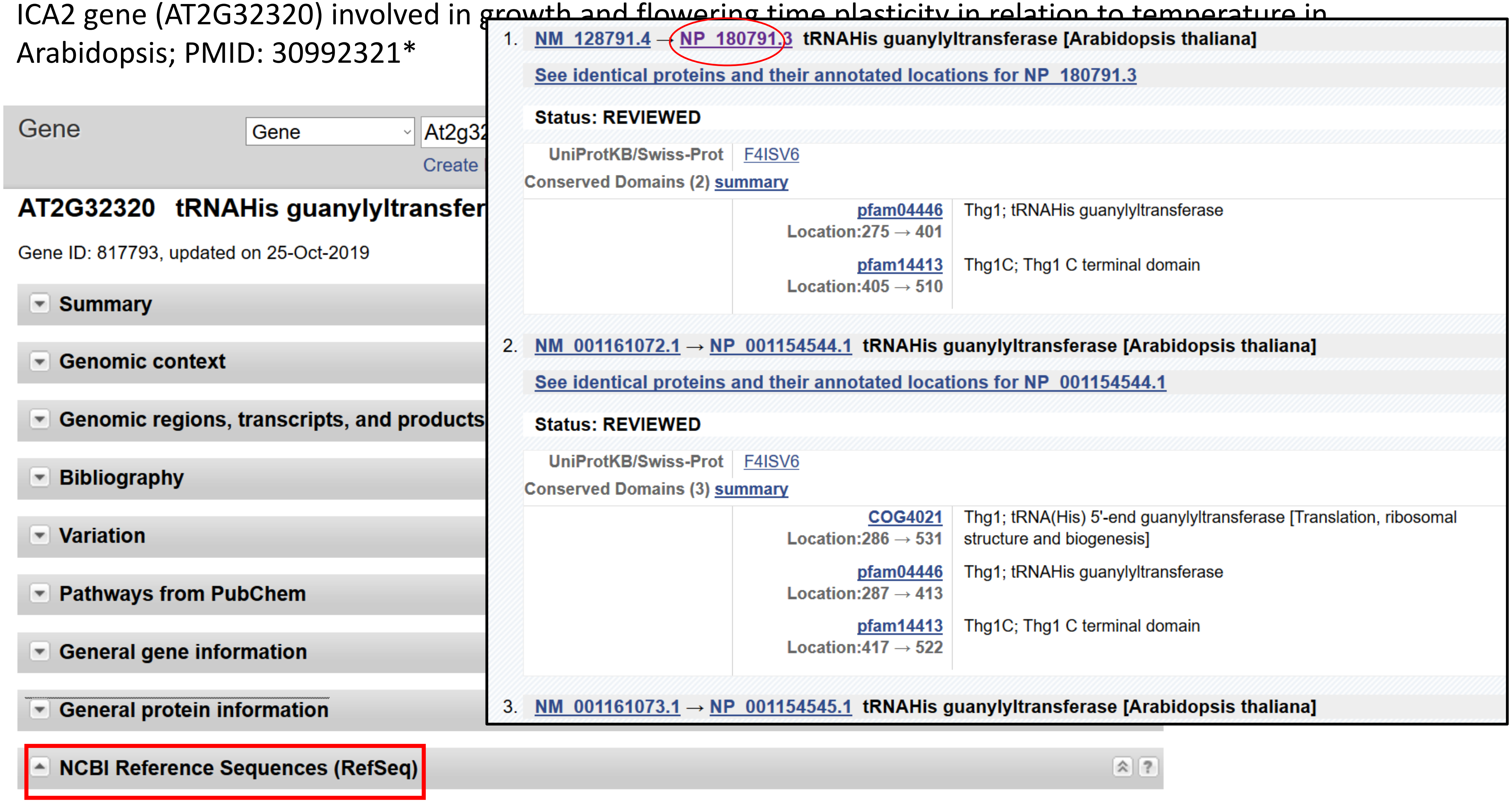

GLYMA_01G162600
mRNA-hypothetical ...
KRH76596.1
- RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 10
310
310
- RNA-seq intron features, aggregate (filtered, unique hits), NCBI Glycine max Annotati

*Seihl et. al. Plant Physiol. 2014 Nov;166(3):1162-1176. Broad 4-hydroxyphenylpyruvate dioxygenase inhibitor herbicide tolerance in soybean with an optimized enzyme and expression cassette

**Liew et. Al. Sci Rep. 2017 Sep 6;7(1):10605. A novel role of the soybean clock gene LUX ARRHYTHMO in male reproductive development.

Data Analysis using BLAST

ICA2 gene (AT2G32320) involved in growth and flowering time plasticity in relation to temperature in Arabidopsis; PMID: 30992321*



Gene

Gene

At2g32

Create

AT2G32320 tRNAHis guanylyltransfer

Gene ID: 817793, updated on 25-Oct-2019

- **Summary**
- **Genomic context**
- **Genomic regions, transcripts, and products**
- **Bibliography**
- **Variation**
- **Pathways from PubChem**
- **General gene information**
- **General protein information**
- **NCBI Reference Sequences (RefSeq)**

1. NM_128791.4 → NP_180791.3 **tRNAHis guanylyltransferase [Arabidopsis thaliana]**

   See identical proteins and their annotated locations for NP_180791.3

   Status: REVIEWED

   UniProtKB/Swiss-Prot F4ISV6

   Conserved Domains (2) summary

pfam04446 Location:275 → 401	Thg1; tRNAHis guanylyltransferase
pfam14413 Location:405 → 510	Thg1C; Thg1 C terminal domain

2. NM_001161072.1 → NP_001154544.1 **tRNAHis guanylyltransferase [Arabidopsis thaliana]**

   See identical proteins and their annotated locations for NP_001154544.1

   Status: REVIEWED

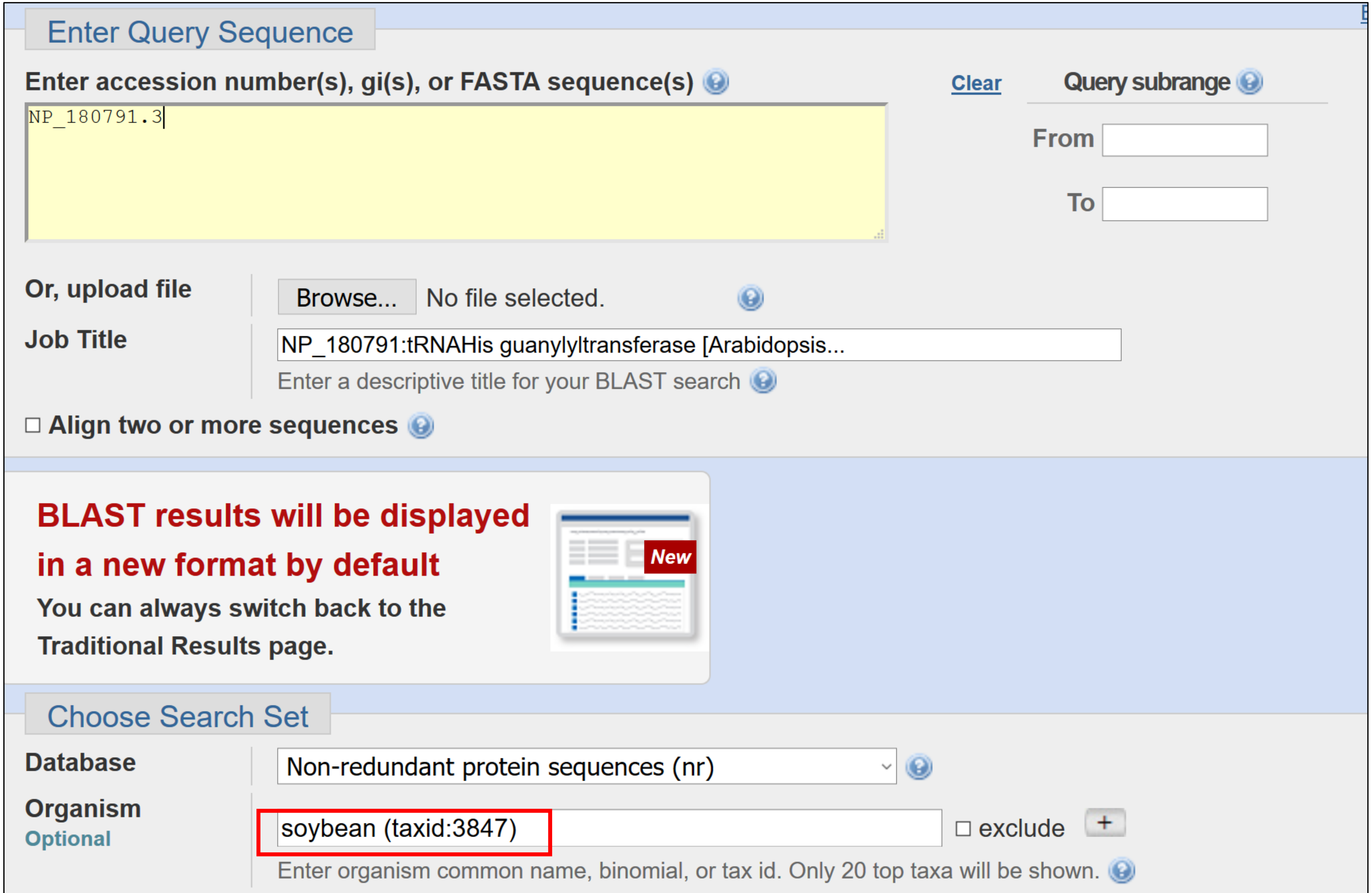
   UniProtKB/Swiss-Prot F4ISV6

   Conserved Domains (3) summary

COG4021 Location:286 → 531	Thg1; tRNA(His) 5'-end guanylyltransferase [Translation, ribosomal structure and biogenesis]
pfam04446 Location:287 → 413	Thg1; tRNAHis guanylyltransferase
pfam14413 Location:417 → 522	Thg1C; Thg1 C terminal domain

3. NM_001161073.1 → NP_001154545.1 **tRNAHis guanylyltransferase [Arabidopsis thaliana]**

*Mendez-Vigo et. al. Plant Cell. 2019 Jun;31(6):1222-1237. Genetic Interactions and Molecular Evolution of the Duplicated Genes *ICARUS2* and *ICARUS1* Help Arabidopsis Plants Adapt to Different Ambient Temperatures.



Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

NP_180791.3

Clear

Query subrange

From

To

Or, upload file: Browse... No file selected.

Job Title: NP_180791:tRNAHis guanylyltransferase [Arabidopsis...

Enter a descriptive title for your BLAST search

□ Align two or more sequences

BLAST results will be displayed in a new format by default

You can always switch back to the Traditional Results page.

Choose Search Set

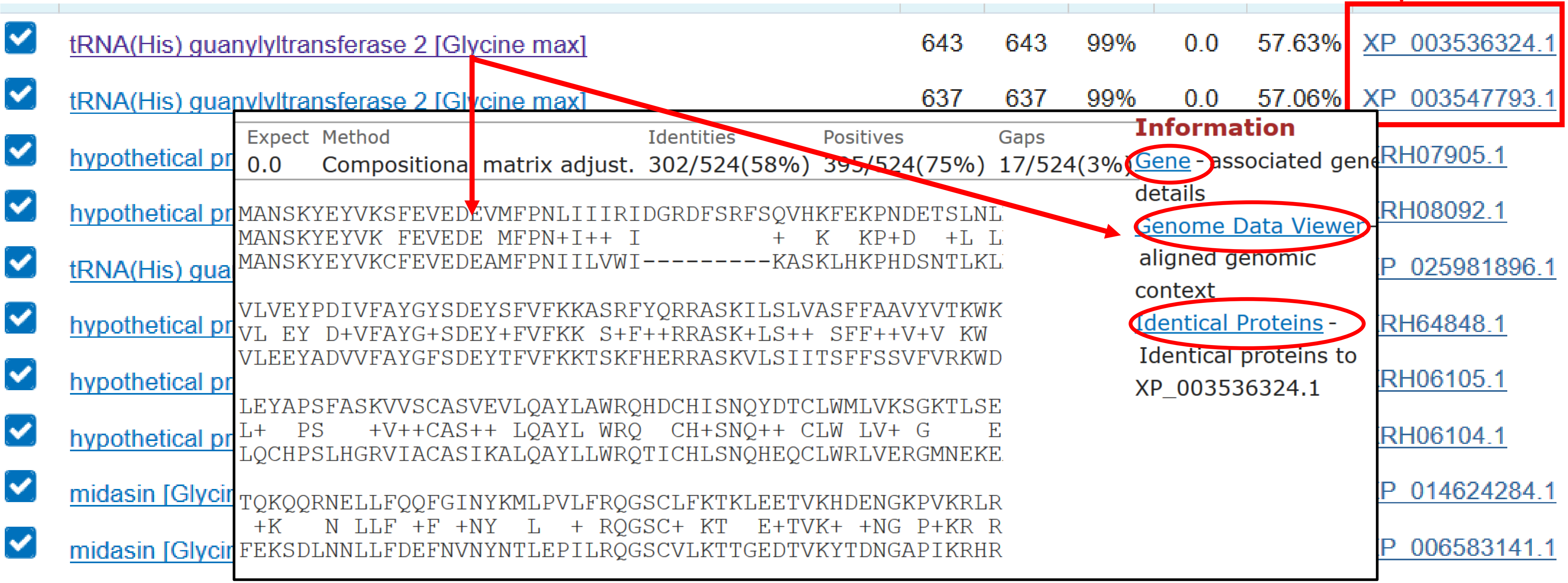
Database: Non-redundant protein sequences (nr)

Organism Optional: soybean (taxid:3847) □ exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Accessing RefSeq from BLAST output

Select RefSeqs



☑	tRNA(His) guanylyltransferase 2 [Glycine max]	643	643	99%	0.0	57.63%	XP_003536324.1
☑	tRNA(His) guanylyltransferase 2 [Glycine max]	637	637	99%	0.0	57.06%	XP_003547793.1
☑	hypothetical pr						RH07905.1
☑	hypothetical pr						RH08092.1
☑	tRNA(His) gua						P_025981896.1
☑	hypothetical pr						RH64848.1
☑	hypothetical pr						RH06105.1
☑	hypothetical pr						RH06104.1
☑	midasin [Glycir						P_014624284.1
☑	midasin [Glycir						P_006583141.1

Expect	Method	Identities	Positives	Gaps
0.0	Compositiona matrix adjust.	302/524(58%)	395/524(75%)	17/524(3%)

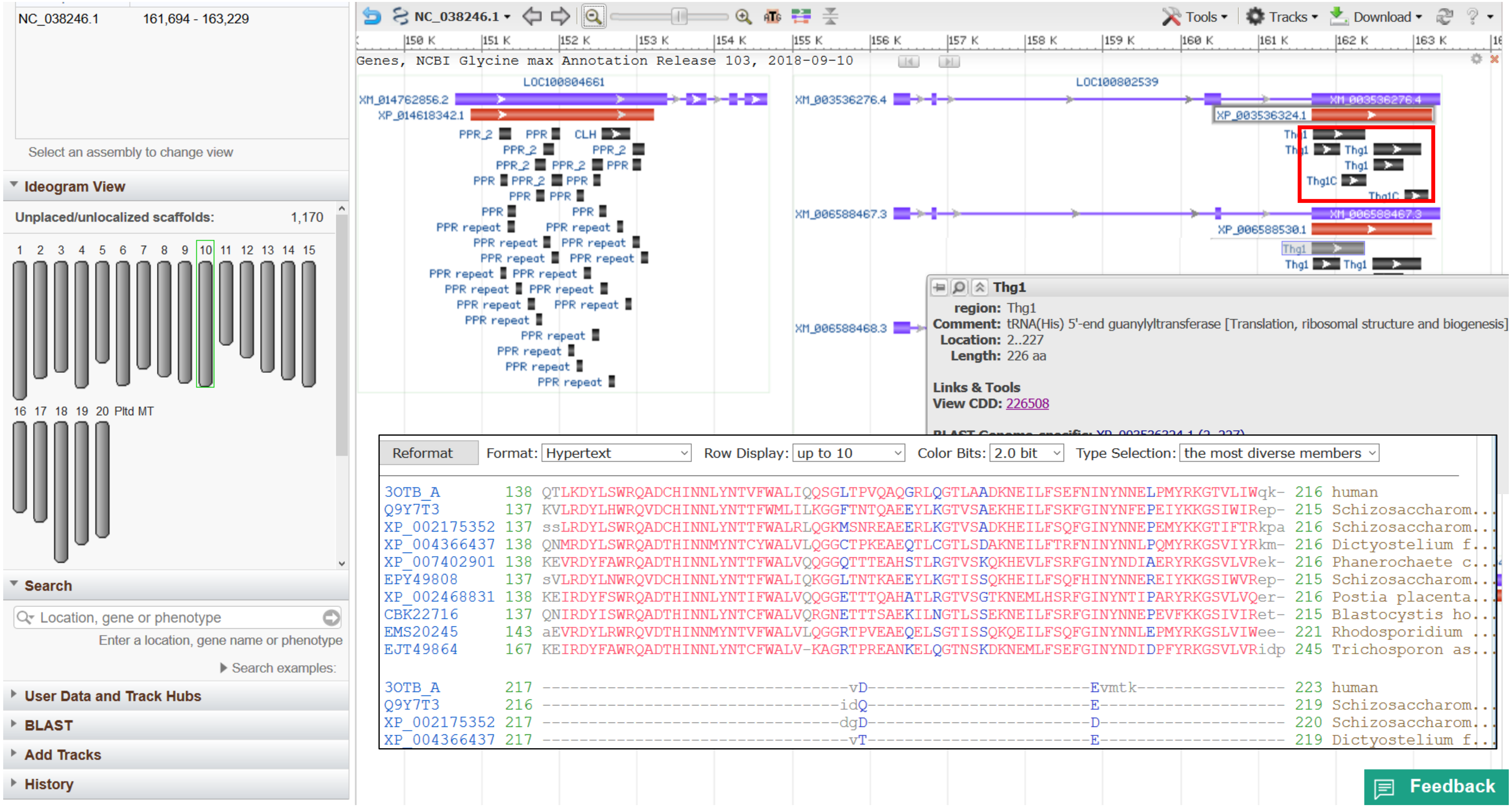
```
MANSKYEYVKSFEVEDEVMFPNLIIIRIDGRDFSRFSQVHKFEKPNDETSLNL
MANSKYEYVK FEVEDE MFPN+I++ I          +  K  KP+D  +L L
MANSKYEYVKCFEVEDEAMFPNIILVWI---------KASKLHKPHDSNTLKL

VLVEYPDIVFAYGYSDEYSFVFKKASRFYQRRASKILSLVASFFAAVYVTKWK
VL EY D+VFAYG+SDEY+FVFKK S+F++RRASK+LS++ SFF++V+V KW
VLEEYADVVFAYGFSDEYTFVFKKTSKFHERRASKVLSIITSFFSSVFVRKWD

LEYAPSFASKVVSCASVEVLQAYLAWRQHDCHISNQYDTCLWMLVKSGKTLSE
L+  PS   +V++CAS++ LQAYL WRQ  CH+SNQ++ CLW LV+ G     E
LQCHPSLHGRVIACASIKALQAYLLWRQTICHLSNQHEQCLWRLVERGMNEKE

TQKQQRNELLFQQFGINYKMLPVLFRQGSCLFKTKLEETVKHDENGKPVKRLR
 +K   N LLF +F +NY   L  + RQGSC+ KT  E+TVK+ +NG P+KR R
FEKSDLNNLLFDEFNVNYNTLEPILRQGSCVLKTTGEDTVKYTDNGAPIKRHR
```

Information

Gene - associated gene details

Genome Data Viewer - aligned genomic context

Identical Proteins - Identical proteins to XP_003536324.1



Thg1

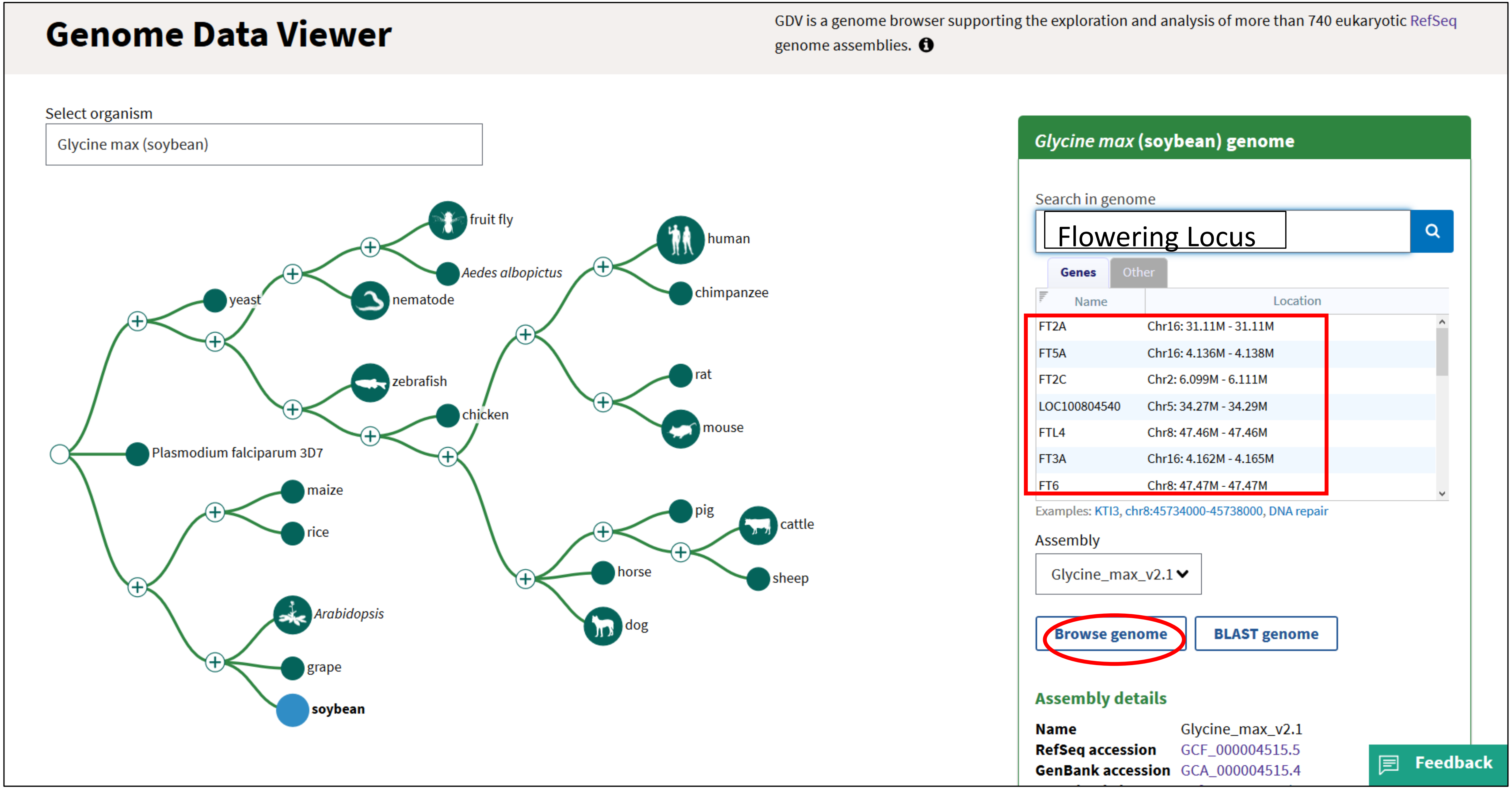
region: Thg1
Comment: tRNA(His) 5'-end guanylyltransferase [Translation, ribosomal structure and biogenesis]
Location: 2..227
Length: 226 aa

Links & Tools
View CDD: 226508

BLAST Genome-specific: XP_003536324.1 (2..227)

Reformat Format: Hypertext Row Display: up to 10 Color Bits: 2.0 bit Type Selection: the most diverse members

```
3OTB_A       138 QTLKDYLSWRQADCHINNLYNTVFWALIQQSGLTPVQAQGRLQGTLAADKNEILFSEFNINYNNELPMYRKGTVLIWqk- 216 human
Q9Y7T3       137 KVLRDYLHWRQVDCHINNLYNTTFWMLILKGGFTNTQAEEYLKGTVSAEKHEILFSKFGINYNFEPEIYKKGSIWIRep- 215 Schizosaccharom..
XP_002175352 137 ssLRDYLSWRQADCHINNLYNTTFWALRLQGKMSNREAEERLKGTVSADKHEILFSQFGINYNNEPEMYKKGTIFTRkpa 216 Schizosaccharom..
XP_004366437 138 QNMRDYLSWRQADTHINNMYNTCYWALVLQGGCTPKEAEQTLCGTLSDAKNEILFTRFNINYNNLPQMYRKGSVIYRkm- 216 Dictyostelium f..
XP_007402901 138 KEVRDYFAWRQADTHINNLYNTTFWALVQQGGQTTTEAHSTLRGTVSKQKHEVLFSRFGINYNDIAERYRKGSVLVRek- 216 Phanerochaete c..
EPY49808     137 sVLRDYLNWRQVDCHINNLYNTTFWALIQKGGLTNTKAEEYLKGTISSQKHEILFSQFHINYNNEREIYKKGSIWVRep- 215 Schizosaccharom..
XP_002468831 138 KEIRDYFSWRQADTHINNLYNTIFWALVQQGGETTTQAHATLRGTVSGTKNEMLHSRFGINYNTIPARYRKGSVLVQer- 216 Postia placenta..
CBK22716     137 QNIRDYISWRQADTHINNLYNTCFWALVQRGNETTTSAEKILNGTLSSEKNEILFSRFGINYNNEPEVFKKGSIVIRet- 215 Blastocystis ho..
EMS20245     143 aEVRDYLRWRQVDTHINNMYNTVFWALVLQGGRTPVEAEQELSGTISSQKQEILFSQFGINYNNLEPMYRKGSLVIWee- 221 Rhodosporidium ..
EJT49864     167 KEIRDYFAWRQADTHINNLYNTCFWALV-KAGRTPREANKELQGTNSKDKNEMLFSEFGINYNDIDPFYRKGSVLVRidp 245 Trichosporon as..

3OTB_A       217 ---------------------------------vD------------------------Evmtk---------------- 223 human
Q9Y7T3       216 --------------------------------idQ------------------------E-------------------- 219 Schizosaccharom..
XP_002175352 217 --------------------------------dgD------------------------D-------------------- 220 Schizosaccharom..
XP_004366437 217 ---------------------------------vT------------------------E-------------------- 219 Dictyostelium f..
```



Genome Data Viewer

GDV is a genome browser supporting the exploration and analysis of more than 740 eukaryotic RefSeq genome assemblies.

Select organism

Glycine max (soybean)

Glycine max (soybean) genome

Search in genome

Flowering Locus

Name	Location
FT2A	Chr16: 31.11M - 31.11M
FT5A	Chr16: 4.136M - 4.138M
FT2C	Chr2: 6.099M - 6.111M
LOC100804540	Chr5: 34.27M - 34.29M
FTL4	Chr8: 47.46M - 47.46M
FT3A	Chr16: 4.162M - 4.165M
FT6	Chr8: 47.47M - 47.47M

Examples: KTI3, chr8:45734000-45738000, DNA repair

Assembly

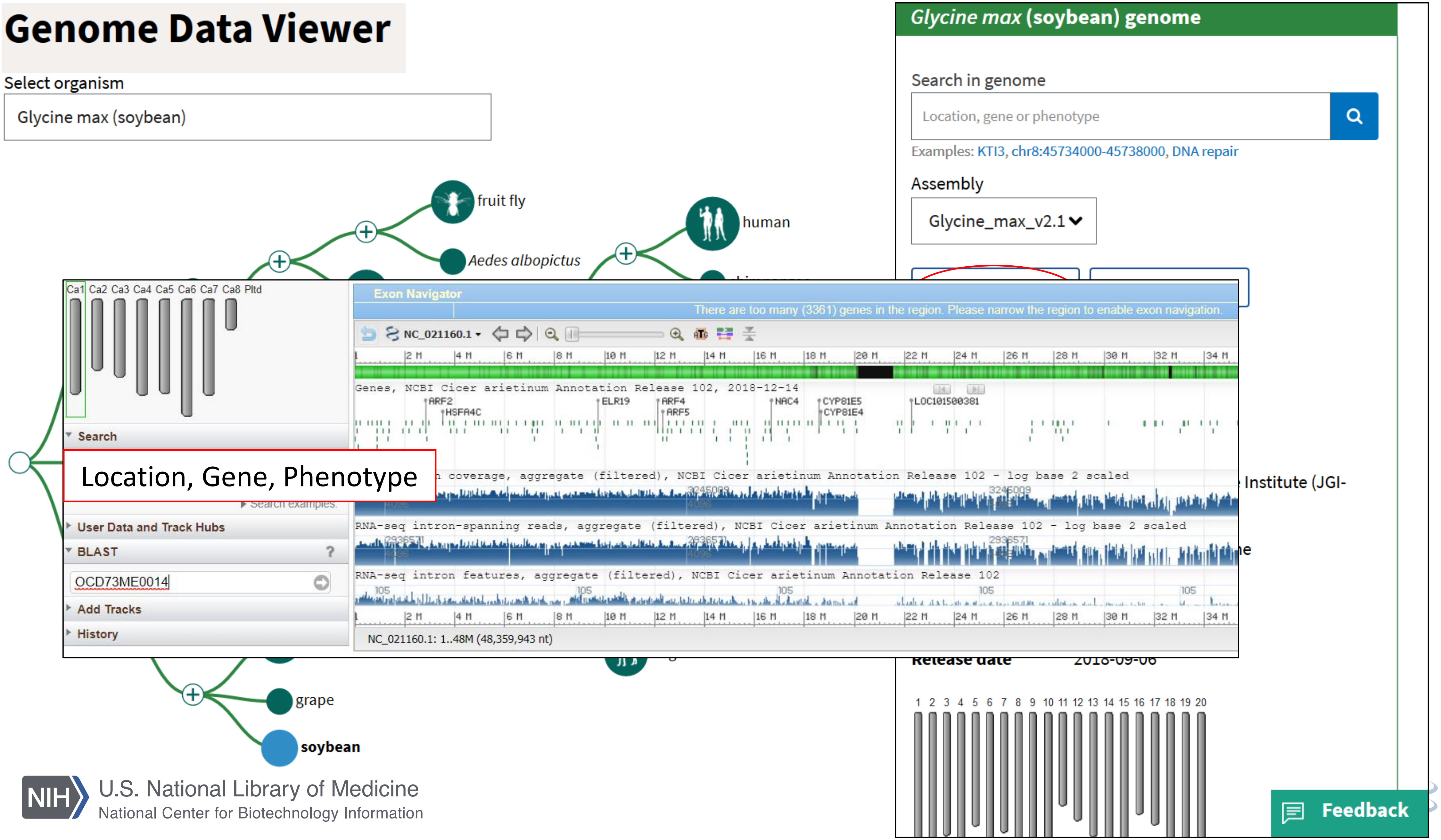
Glycine_max_v2.1

Assembly details

Name Glycine_max_v2.1

RefSeq accession GCF_000004515.5

GenBank accession GCA_000004515.4



Genome Data Viewer

Select organism

Glycine max (soybean)

Glycine max (soybean) genome

Search in genome

Location, gene or phenotype

Examples: KTI3, chr8:45734000-45738000, DNA repair

Assembly

Glycine_max_v2.1

Institute (JGI-

Release date 2018-09-06

Location, Gene, Phenotype

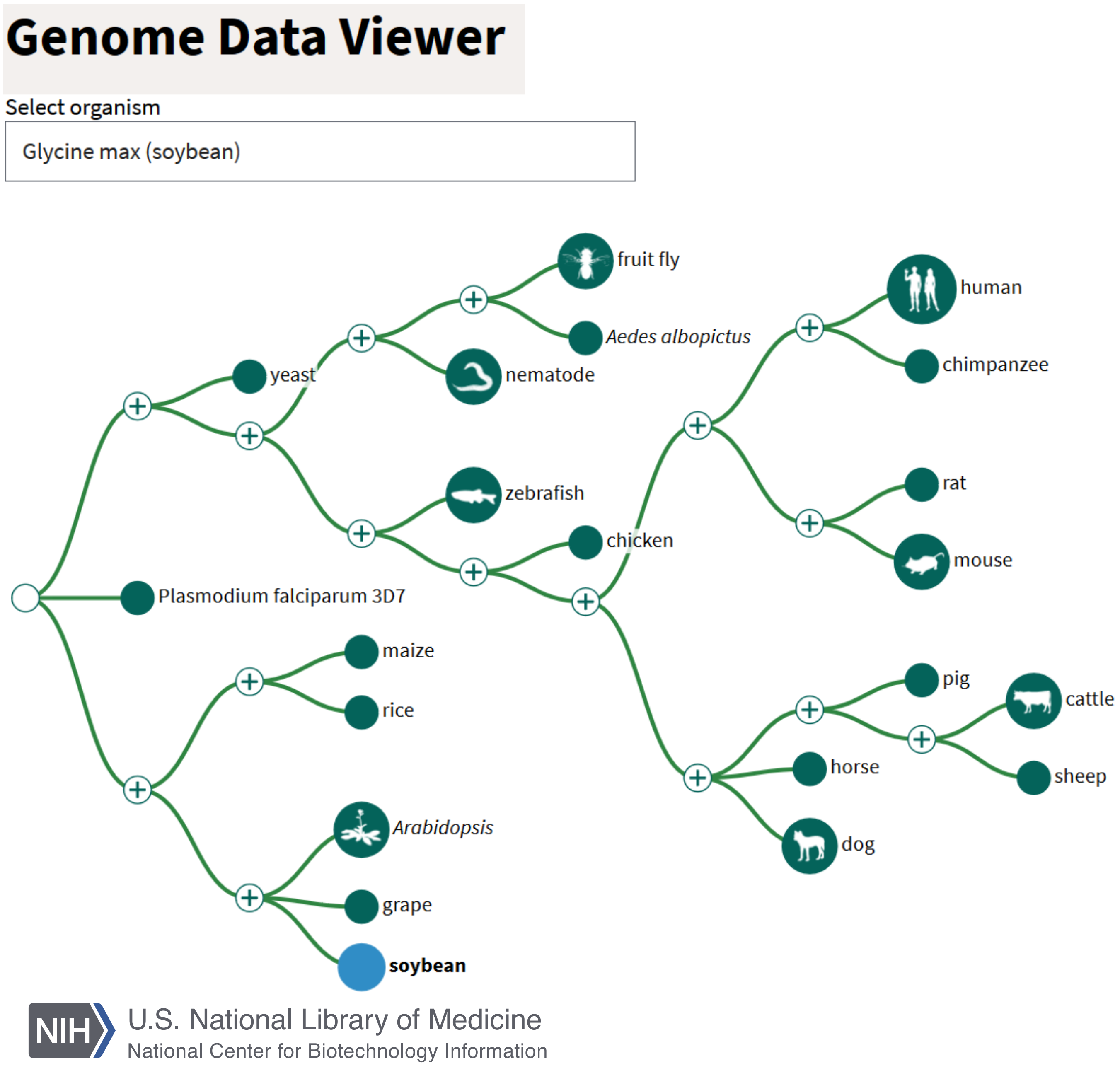
OCD73ME0014

NC_021160.1: 1..48M (48,359,943 nt)

There are too many (3361) genes in the region. Please narrow the region to enable exon navigation.

fruit fly
Aedes albopictus
human
grape
soybean
Ca1
Ca2
Ca3
Ca4
Ca5
Ca6
Ca7
Ca8
Pltd
Search
Search examples
User Data and Track Hubs
BLAST
?
Add Tracks
History
Exon Navigator
NC_021160.1
2 M
4 M
6 M
8 M
10 M
12 M
14 M
16 M
18 M
20 M
22 M
24 M
26 M
28 M
30 M
32 M
34 M
Genes, NCBI Cicer arietinum Annotation Release 102, 2018-12-14
ARF2
HSFA4C
ELR19
ARF4
ARF5
NAC4
CYP81E5
CYP81E4
LOC101500381
coverage, aggregate (filtered), NCBI Cicer arietinum Annotation Release 102 - log base 2 scaled
3245009
4096
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Cicer arietinum Annotation Release 102 - log base 2 scaled
2936571
RNA-seq intron features, aggregate (filtered), NCBI Cicer arietinum Annotation Release 102
105
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Feedback

U.S. National Library of Medicine
National Center for Biotechnology Information

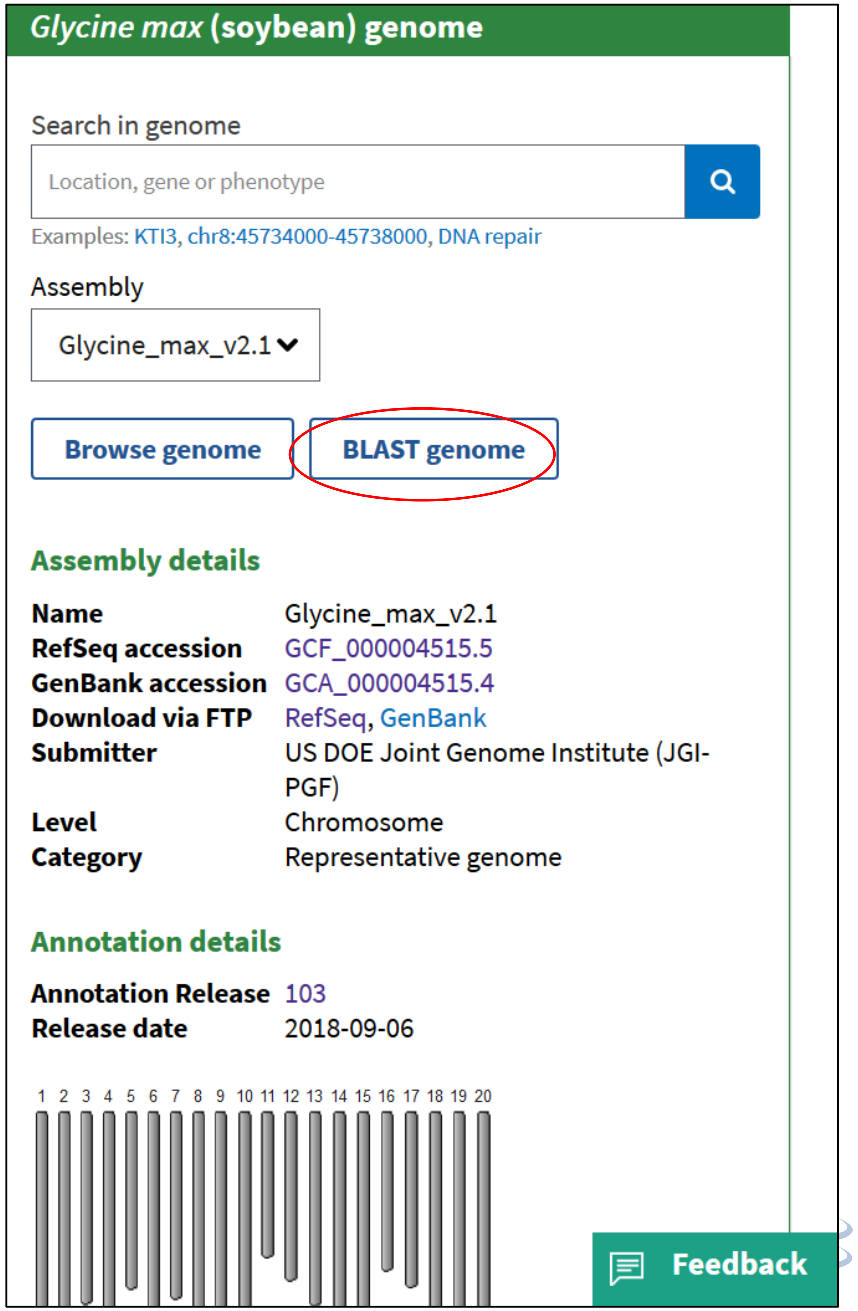


Genome Data Viewer

Select organism

Glycine max (soybean)

U.S. National Library of Medicine
National Center for Biotechnology Information



Glycine max (soybean) genome

Search in genome

Location, gene or phenotype

Examples: KTI3, chr8:45734000-45738000, DNA repair

Assembly

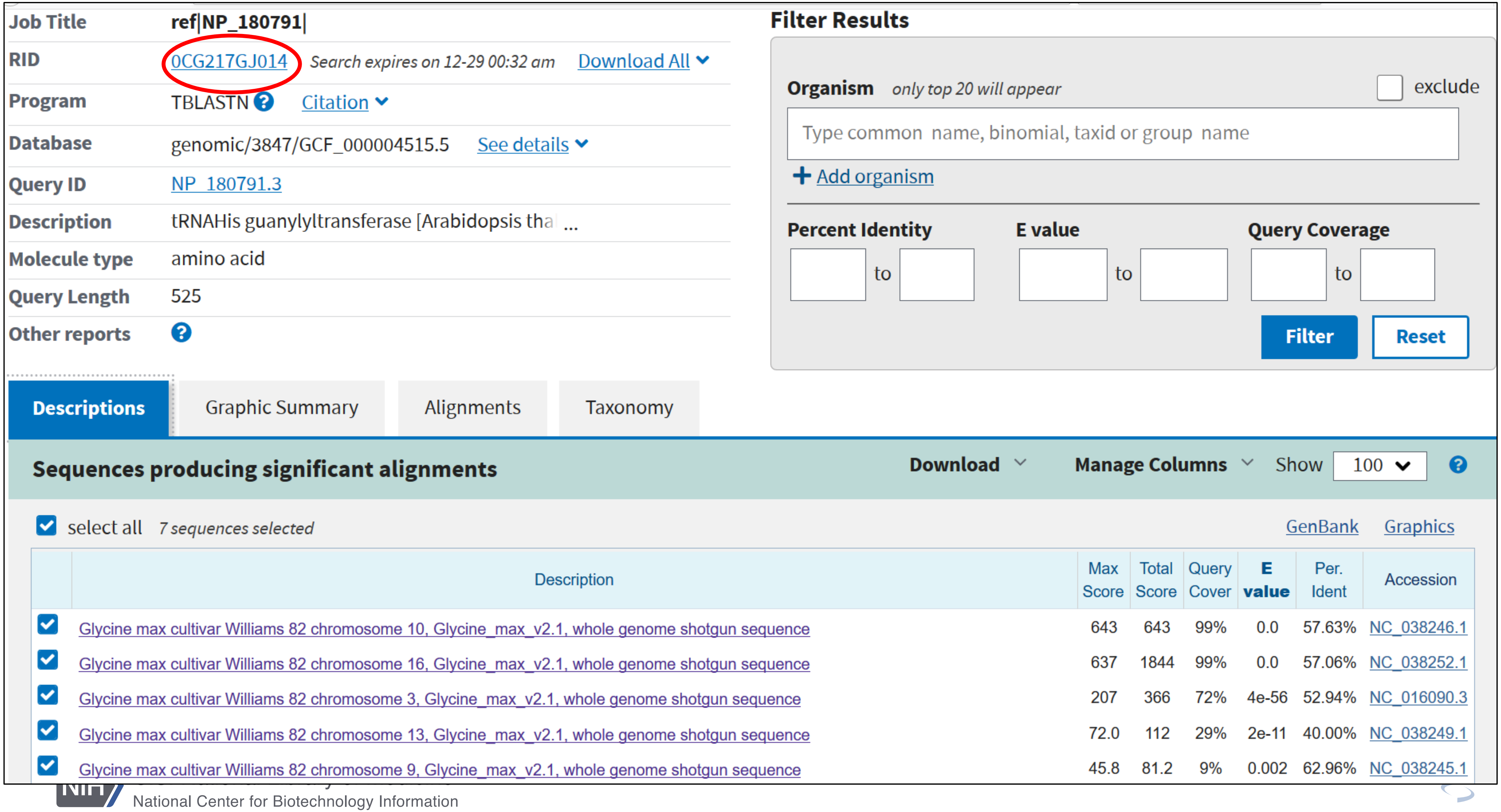
Glycine_max_v2.1

Assembly details

Name	Glycine_max_v2.1
RefSeq accession	GCF_000004515.5
GenBank accession	GCA_000004515.4
Download via FTP	RefSeq, GenBank
Submitter	US DOE Joint Genome Institute (JGI-PGF)
Level	Chromosome
Category	Representative genome

Annotation details

Annotation Release	103
Release date	2018-09-06



Job Title	ref|NP_180791|
RID	0CG217GJ014 *Search expires on 12-29 00:32 am* Download All
Program	TBLASTN ? Citation
Database	genomic/3847/GCF_000004515.5 See details
Query ID	NP_180791.3
Description	tRNAHis guanylyltransferase [Arabidopsis tha ...
Molecule type	amino acid
Query Length	525
Other reports	?

Filter Results

Organism *only top 20 will appear* □ exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to **E value** to **Query Coverage** to

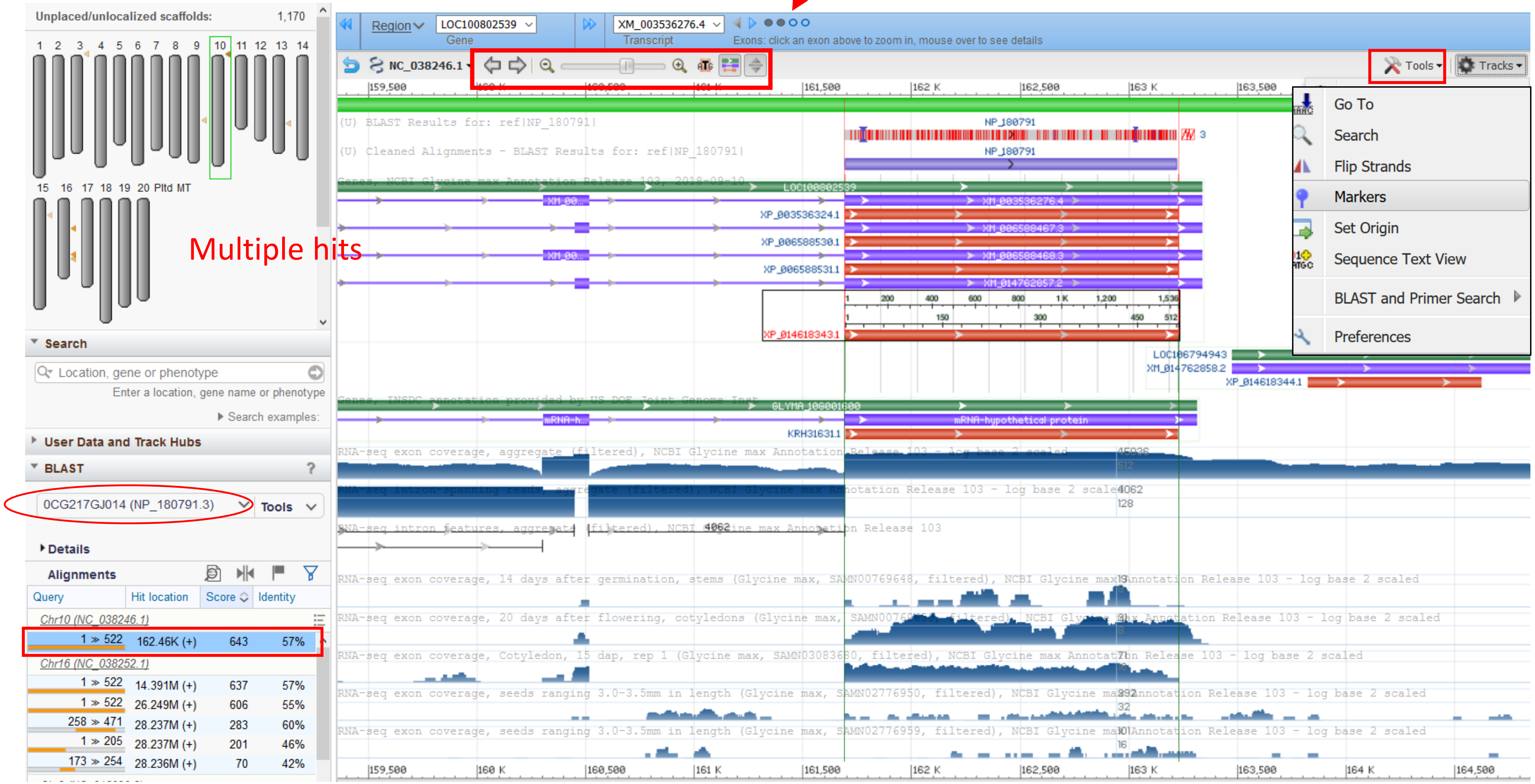
Filter Reset

Sequences producing significant alignments

☑ select all *7 sequences selected*

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	Glycine max cultivar Williams 82 chromosome 10, Glycine_max_v2.1, whole genome shotgun sequence	643	643	99%	0.0	57.63%	NC_038246.1
☑	Glycine max cultivar Williams 82 chromosome 16, Glycine_max_v2.1, whole genome shotgun sequence	637	1844	99%	0.0	57.06%	NC_038252.1
☑	Glycine max cultivar Williams 82 chromosome 3, Glycine_max_v2.1, whole genome shotgun sequence	207	366	72%	4e-56	52.94%	NC_016090.3
☑	Glycine max cultivar Williams 82 chromosome 13, Glycine_max_v2.1, whole genome shotgun sequence	72.0	112	29%	2e-11	40.00%	NC_038249.1
☑	Glycine max cultivar Williams 82 chromosome 9, Glycine_max_v2.1, whole genome shotgun sequence	45.8	81.2	9%	0.002	62.96%	NC_038245.1

Pick the exon

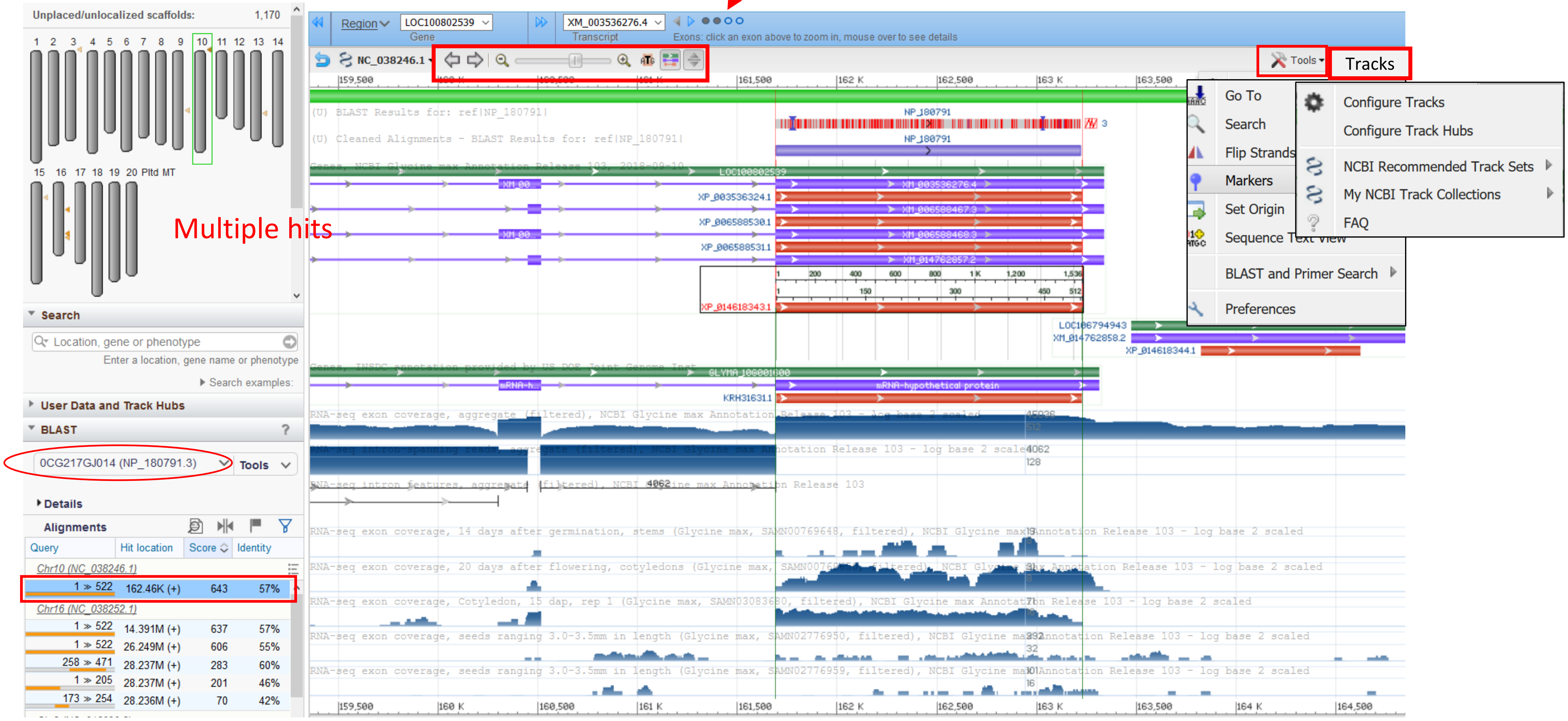


Multiple hits

Unplaced/unlocalized scaffolds: 1,170
1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 Pltd MT
Search
Location, gene or phenotype
Enter a location, gene name or phenotype
Search examples:
User Data and Track Hubs
BLAST
0CG217GJ014 (NP_180791.3)
Tools
Details
Alignments
Region
LOC100802539
Gene
XM_003536276.4
Transcript
Exons: click an exon above to zoom in, mouse over to see details
NC_038246.1
Tools
Tracks
Go To
Search
Flip Strands
Markers
Set Origin
Sequence Text View
BLAST and Primer Search
Preferences
159,500
160 K
160,500
161 K
161,500
162 K
162,500
163 K
163,500
164 K
164,500
(U) BLAST Results for: ref|NP_180791|
NP_180791
(U) Cleaned Alignments - BLAST Results for: ref|NP_180791|
NP_180791
Genes, NCBI Glycine max Annotation Release 103, 2018-09-10
LOC100802539
XM_003536276.4
XP_003536324.1
XM_006588467.3
XP_006588530.1
XM_006588468.3
XP_006588531.1
XM_014762857.2
XP_014618343.1
LOC106794943
XM_014762858.2
XP_014618344.1
Genes, INSDC annotation provided by US DOE Joint Genome Inst
GLYMA_10G001600
mRNA-hypothetical protein
KRH31631.1
RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq intron features, aggregate (filtered), NCBI Glycine max Annotation Release 103
RNA-seq exon coverage, 14 days after germination, stems (Glycine max, SAMN00769648, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine max, SAMN00769..., filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, Cotyledon, 15 dap, rep 1 (Glycine max, SAMN03083680, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max, SAMN02776950, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max, SAMN02776959, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

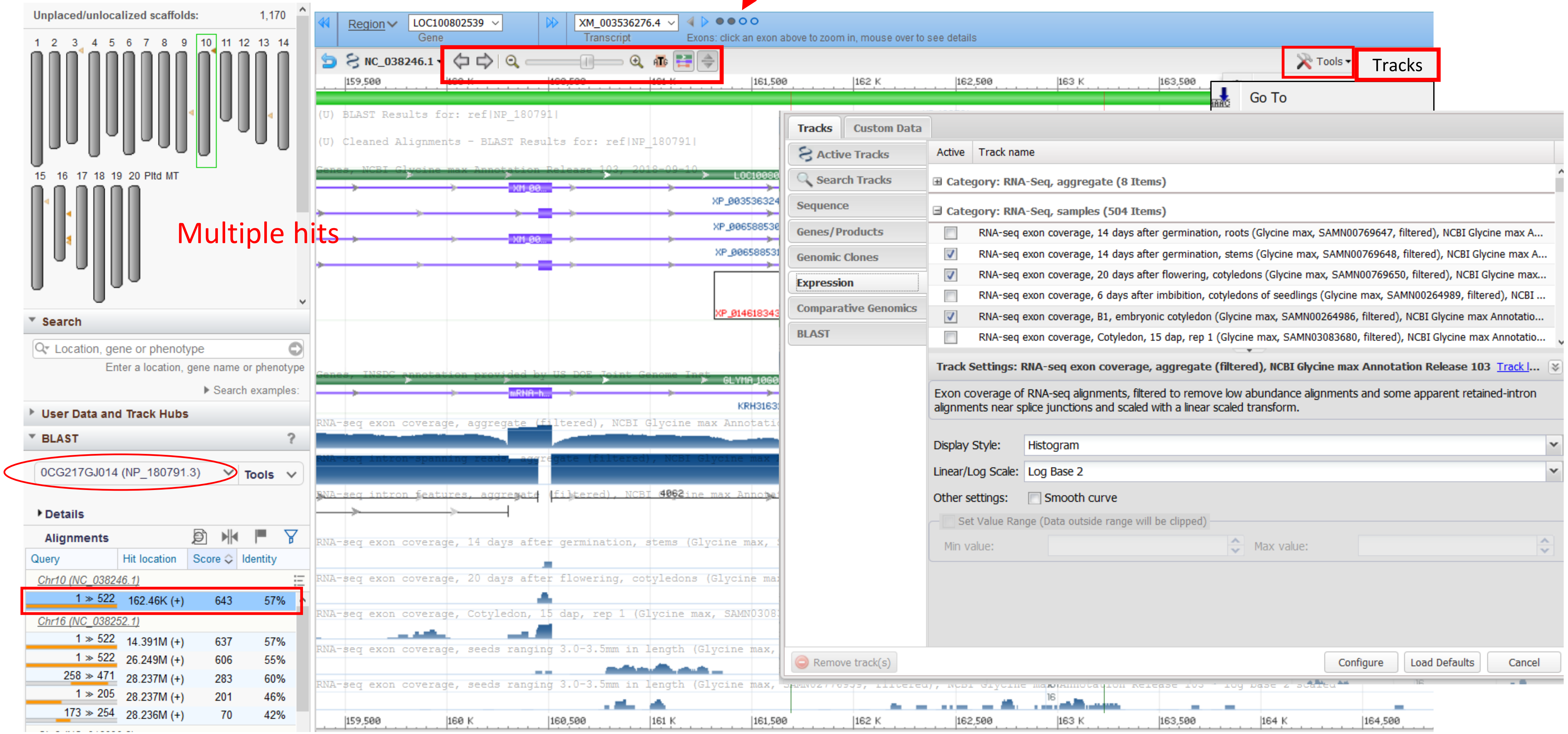
Query	Hit location	Score	Identity
Chr10 (NC_038246.1)			
1 » 522	162.46K (+)	643	57%
Chr16 (NC_038252.1)			
1 » 522	14.391M (+)	637	57%
1 » 522	26.249M (+)	606	55%
258 » 471	28.237M (+)	283	60%
1 » 205	28.237M (+)	201	46%
173 » 254	28.236M (+)	70	42%

Pick the exon


Unplaced/unlocalized scaffolds: 1,170
1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 Pltd MT
Multiple hits
Search
Location, gene or phenotype
Enter a location, gene name or phenotype
Search examples:
User Data and Track Hubs
BLAST
0CG217GJ014 (NP_180791.3)
Tools
Details
Alignments
Region
LOC100802539
Gene
XM_003536276.4
Transcript
Exons: click an exon above to zoom in, mouse over to see details
NC_038246.1
Tools
Tracks
Go To
Search
Flip Strands
Markers
Set Origin
Sequence Text View
BLAST and Primer Search
Preferences
Configure Tracks
Configure Track Hubs
NCBI Recommended Track Sets
My NCBI Track Collections
FAQ
159,500
160 K
160,500
161 K
161,500
162 K
162,500
163 K
163,500
164 K
164,500
(U) BLAST Results for: ref|NP_180791|
NP_180791
(U) Cleaned Alignments - BLAST Results for: ref|NP_180791|
Genes, NCBI Glycine max Annotation Release 103, 2018-09-10
LOC100802539
XM_003536276.4
XP_003536324.1
XM_006588467.3
XP_006588530.1
XM_006588468.3
XP_006588531.1
XM_014762857.2
XP_014618343.1
LOC106794943
XM_014762858.2
XP_014618344.1
Genes, INSDC annotation provided by US DOE Joint Genome Inst
GLYMA_10G001600
mRNA-hypothetical protein
KRH31631.1
RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq intron features, aggregate (filtered), NCBI Glycine max Annotation Release 103
RNA-seq exon coverage, 14 days after germination, stems (Glycine max, SAMN00769648, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine max, SAMN00769..., filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, Cotyledon, 15 dap, rep 1 (Glycine max, SAMN03083680, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max, SAMN02776950, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max, SAMN02776959, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Query	Hit location	Score	Identity
Chr10 (NC_038246.1)			
1 » 522	162.46K (+)	643	57%
Chr16 (NC_038252.1)			
1 » 522	14.391M (+)	637	57%
1 » 522	26.249M (+)	606	55%
258 » 471	28.237M (+)	283	60%
1 » 205	28.237M (+)	201	46%
173 » 254	28.236M (+)	70	42%

Pick the exon



Multiple hits

Tracks

Go To

Unplaced/unlocalized scaffolds: 1,170

Search

Location, gene or phenotype

Enter a location, gene name or phenotype

Search examples:

User Data and Track Hubs

BLAST

0CG217GJ014 (NP_180791.3) Tools

Details

Alignments

Query	Hit location	Score	Identity
Chr10 (NC_038246.1)			
1 » 522	162.46K (+)	643	57%
Chr16 (NC_038252.1)			
1 » 522	14.391M (+)	637	57%
1 » 522	26.249M (+)	606	55%
258 » 471	28.237M (+)	283	60%
1 » 205	28.237M (+)	201	46%
173 » 254	28.236M (+)	70	42%

Region LOC100802539 Gene XM_003536276.4 Transcript

Exons: click an exon above to zoom in, mouse over to see details

NC_038246.1

(U) BLAST Results for: ref|NP_180791|

(U) Cleaned Alignments - BLAST Results for: ref|NP_180791|

Genes, NCBI Glycine max Annotation Release 103, 2018-09-10

Genes, INSDC annotation provided by US DOE Joint Genome Inst

RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotati

RNA-seq intron-spanning reads, aggregate (filtered), NCBI Glycine max

RNA-seq intron features, aggregate (filtered), NCBI Glycine max Annota

RNA-seq exon coverage, 14 days after germination, stems (Glycine max,

RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine ma

RNA-seq exon coverage, Cotyledon, 15 dap, rep 1 (Glycine max, SAMN0308

RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max,

RNA-seq exon coverage, seeds ranging 3.0-3.5mm in length (Glycine max, SAMN02776939, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Tools

Tracks | Custom Data

- Active Tracks
- Search Tracks
- Sequence
- Genes/Products
- Genomic Clones
- Expression
- Comparative Genomics
- BLAST

Active	Track name
⊞ Category: RNA-Seq, aggregate (8 Items)	
⊟ Category: RNA-Seq, samples (504 Items)	
☐	RNA-seq exon coverage, 14 days after germination, roots (Glycine max, SAMN00769647, filtered), NCBI Glycine max A...
☑	RNA-seq exon coverage, 14 days after germination, stems (Glycine max, SAMN00769648, filtered), NCBI Glycine max A...
☑	RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine max, SAMN00769650, filtered), NCBI Glycine max...
☐	RNA-seq exon coverage, 6 days after imbibition, cotyledons of seedlings (Glycine max, SAMN00264989, filtered), NCBI ...
☑	RNA-seq exon coverage, B1, embryonic cotyledon (Glycine max, SAMN00264986, filtered), NCBI Glycine max Annotatio...
☐	RNA-seq exon coverage, Cotyledon, 15 dap, rep 1 (Glycine max, SAMN03083680, filtered), NCBI Glycine max Annotatio...

Track Settings: RNA-seq exon coverage, aggregate (filtered), NCBI Glycine max Annotation Release 103 Track l...

Exon coverage of RNA-seq alignments, filtered to remove low abundance alignments and some apparent retained-intron alignments near splice junctions and scaled with a linear scaled transform.

Display Style: Histogram

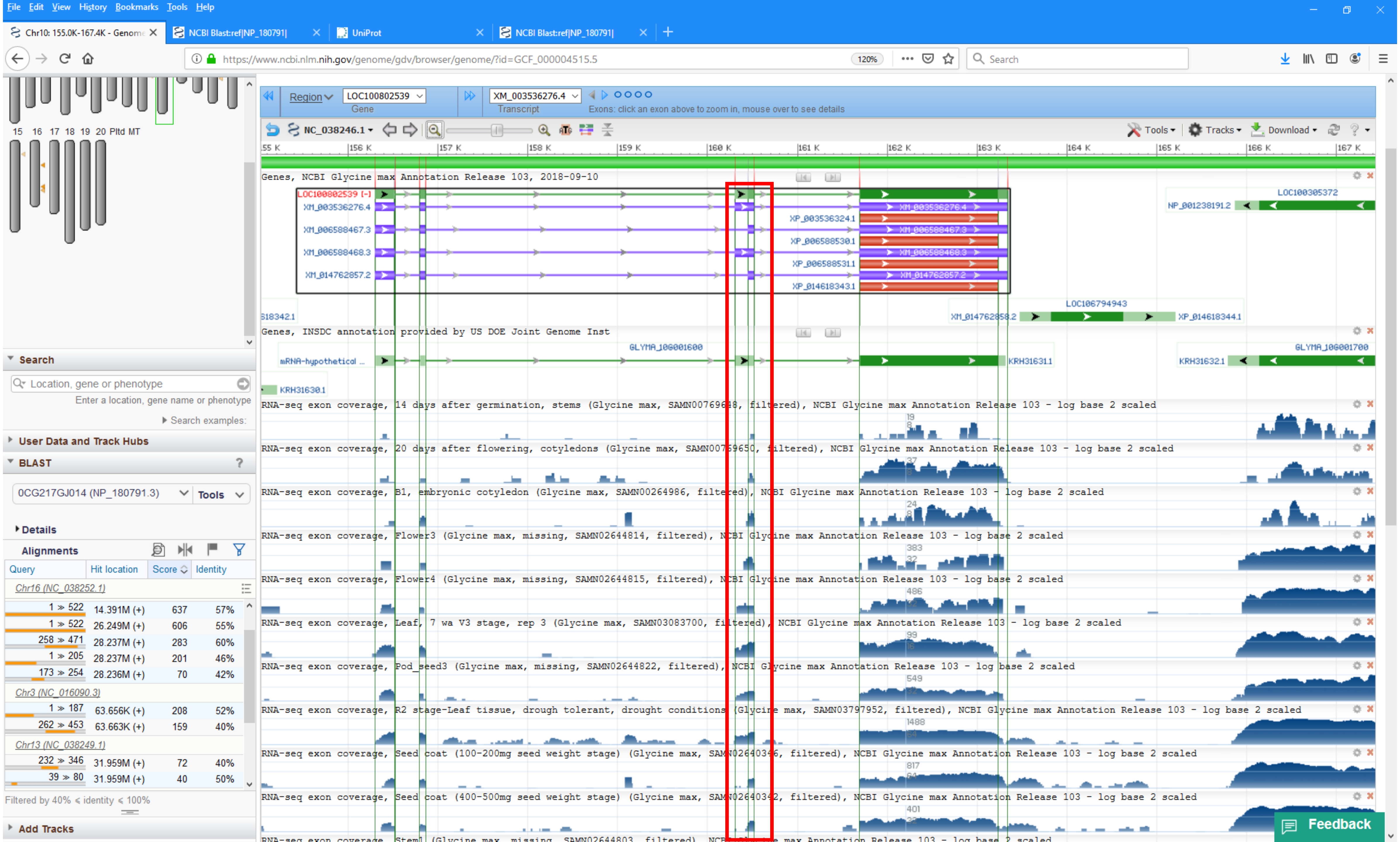
Linear/Log Scale: Log Base 2

Other settings: ☐ Smooth curve

☐ Set Value Range (Data outside range will be clipped)

Min value: Max value:

Remove track(s)   Configure   Load Defaults   Cancel



Search

Location, gene or phenotype

Enter a location, gene name or phenotype

Search examples:

User Data and Track Hubs

BLAST

0CG217GJ014 (NP_180791.3)

Tools

Details

Alignments

Query	Hit location	Score	Identity
Chr16 (NC_038252.1)			
1 ≫ 522	14.391M (+)	637	57%
1 ≫ 522	26.249M (+)	606	55%
258 ≫ 471	28.237M (+)	283	60%
1 ≫ 205	28.237M (+)	201	46%
173 ≫ 254	28.236M (+)	70	42%
Chr3 (NC_016090.3)			
1 ≫ 187	63.656K (+)	208	52%
262 ≫ 453	63.663K (+)	159	40%
Chr13 (NC_038249.1)			
232 ≫ 346	31.959M (+)	72	40%
39 ≫ 80	31.959M (+)	40	50%

Filtered by 40% ≤ identity ≤ 100%

Add Tracks

Genes, NCBI Glycine max Annotation Release 103, 2018-09-10

Genes, INSDC annotation provided by US DOE Joint Genome Inst

RNA-seq exon coverage, 14 days after germination, stems (Glycine max, SAMN00769648, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine max, SAMN00769650, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, B1, embryonic cotyledon (Glycine max, SAMN00264986, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Flower3 (Glycine max, missing, SAMN02644814, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Flower4 (Glycine max, missing, SAMN02644815, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Leaf, 7 wa V3 stage, rep 3 (Glycine max, SAMN03083700, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

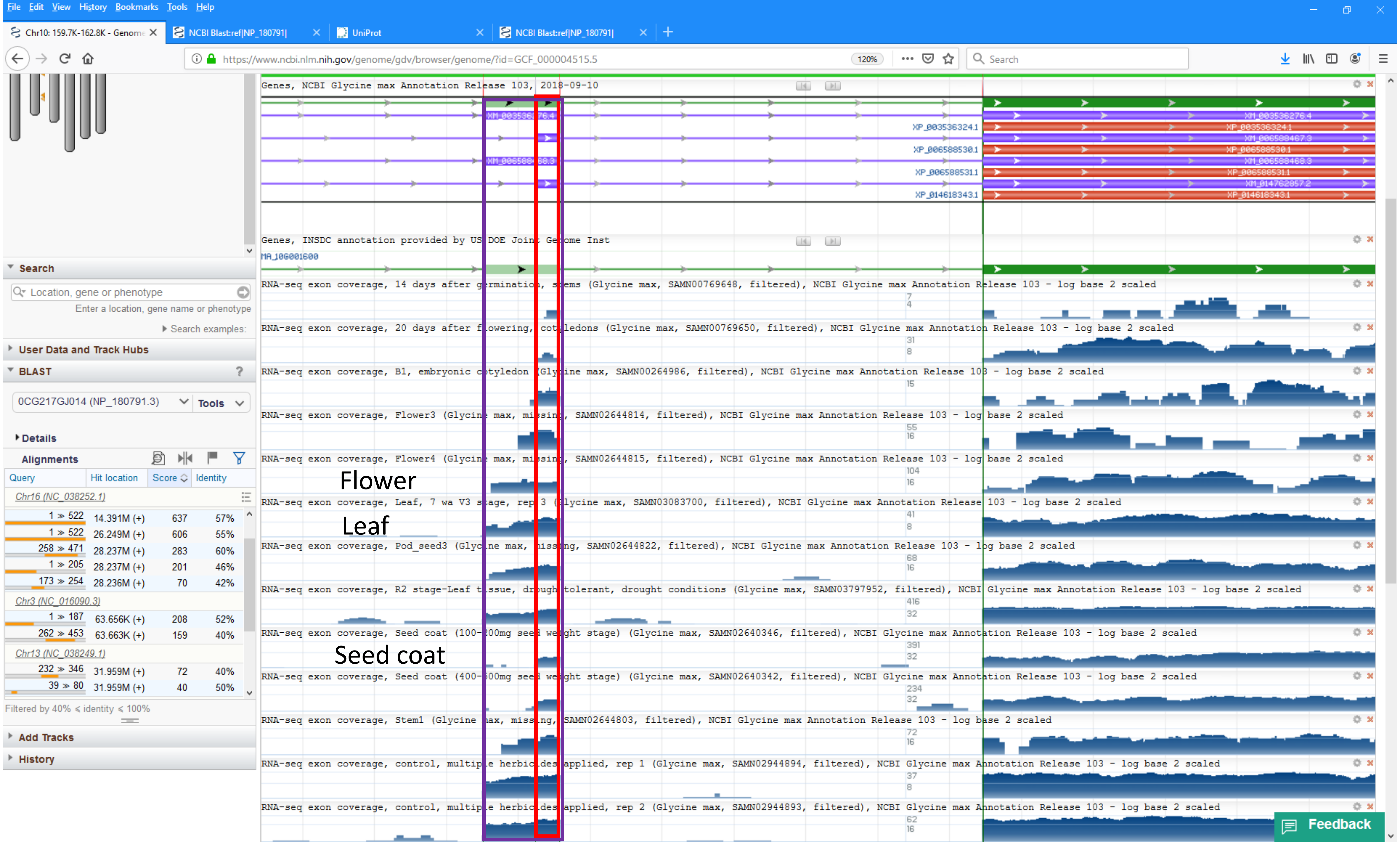
RNA-seq exon coverage, Pod_seed3 (Glycine max, missing, SAMN02644822, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, R2 stage-Leaf tissue, drough tolerant, drought conditions (Glycine max, SAMN03797952, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Seed coat (100-200mg seed weight stage) (Glycine max, SAMN02640346, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Seed coat (400-500mg seed weight stage) (Glycine max, SAMN02640342, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Stem1 (Glycine max, missing, SAMN02644803, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled


File Edit View History Bookmarks Tools Help
Chr10: 159.7K-162.8K - Genome
NCBI Blast:ref|NP_180791|
UniProt
NCBI Blast:ref|NP_180791|
https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000004515.5
120%
Search
XM_003536276.4
XP_003536324.1
XM_006588467.3
XP_006588530.1
XM_006588468.3
XP_006588531.1
XM_014762857.2
XP_014618343.1
MA_106001600
Feedback

Genes, NCBI Glycine max Annotation Release 103, 2018-09-10

Genes, INSDC annotation provided by US DOE Joint Genome Inst

RNA-seq exon coverage, 14 days after germination, stems (Glycine max, SAMN00769648, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, 20 days after flowering, cotyledons (Glycine max, SAMN00769650, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, B1, embryonic cotyledon (Glycine max, SAMN00264986, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Flower3 (Glycine max, missing, SAMN02644814, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Flower4 (Glycine max, missing, SAMN02644815, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Flower

RNA-seq exon coverage, Leaf, 7 wa V3 stage, rep 3 (Glycine max, SAMN03083700, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Leaf

RNA-seq exon coverage, Pod_seed3 (Glycine max, missing, SAMN02644822, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, R2 stage-Leaf tissue, drought tolerant, drought conditions (Glycine max, SAMN03797952, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Seed coat (100-200mg seed weight stage) (Glycine max, SAMN02640346, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Seed coat

RNA-seq exon coverage, Seed coat (400-500mg seed weight stage) (Glycine max, SAMN02640342, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, Stem1 (Glycine max, missing, SAMN02644803, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, control, multiple herbicides applied, rep 1 (Glycine max, SAMN02944894, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

RNA-seq exon coverage, control, multiple herbicides applied, rep 2 (Glycine max, SAMN02944893, filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled

Search

Location, gene or phenotype

Enter a location, gene name or phenotype

Search examples:

User Data and Track Hubs

BLAST

0CG217GJ014 (NP_180791.3) Tools

Details

Alignments

Query	Hit location	Score	Identity
Chr16 (NC_038252.1)			
1 ≫ 522	14.391M (+)	637	57%
1 ≫ 522	26.249M (+)	606	55%
258 ≫ 471	28.237M (+)	283	60%
1 ≫ 205	28.237M (+)	201	46%
173 ≫ 254	28.236M (+)	70	42%
Chr3 (NC_016090.3)			
1 ≫ 187	63.656K (+)	208	52%
262 ≫ 453	63.663K (+)	159	40%
Chr13 (NC_038249.1)			
232 ≫ 346	31.959M (+)	72	40%
39 ≫ 80	31.959M (+)	40	50%

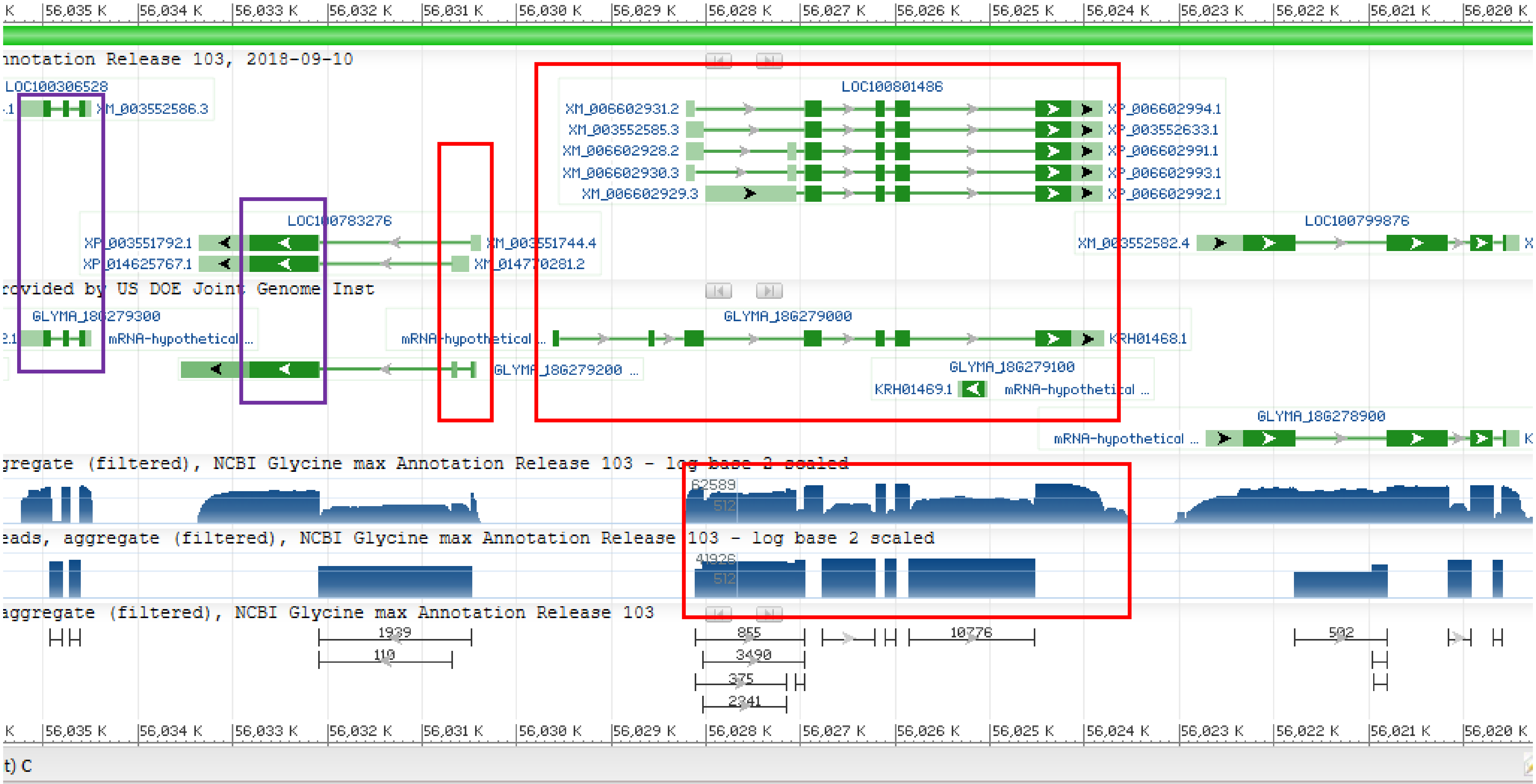
Filtered by 40% ⩽ identity ⩽ 100%

Add Tracks

History

Data Quality

GeneID: 100801486 YIF1B-like protein



K |56,035 K |56,034 K |56,033 K |56,032 K |56,031 K |56,030 K |56,029 K |56,028 K |56,027 K |56,026 K |56,025 K |56,024 K |56,023 K |56,022 K |56,021 K |56,020 K
nnotation Release 103, 2018-09-10
LOC100306528
.1
XM_003552586.3
LOC100801486
XM_006602931.2
XP_006602994.1
XM_003552585.3
XP_003552633.1
XM_006602928.2
XP_006602991.1
XM_006602930.3
XP_006602993.1
XM_006602929.3
XP_006602992.1
LOC100783276
XP_003551792.1
XM_003551744.4
XP_014625767.1
XM_014770281.2
LOC100799876
XM_003552582.4
X
rovided by US DOE Joint Genome Inst
GLYMA_18G279300
2.1
mRNA-hypothetical ...
GLYMA_18G279000
mRNA-hypothetical ...
KRH01468.1
GLYMA_18G279200 ...
GLYMA_18G279100
KRH01469.1
mRNA-hypothetical ...
GLYMA_18G278900
mRNA-hypothetical ...
K
gregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
62589
512
eads, aggregate (filtered), NCBI Glycine max Annotation Release 103 - log base 2 scaled
41926
512
aggregate (filtered), NCBI Glycine max Annotation Release 103
1939
110
855
3490
375
2341
10776
502
K |56,035 K |56,034 K |56,033 K |56,032 K |56,031 K |56,030 K |56,029 K |56,028 K |56,027 K |56,026 K |56,025 K |56,024 K |56,023 K |56,022 K |56,021 K |56,020 K
t) C

*Li et. al., while assessing flavonoid biosynthesis pathway (FBP) genes in various Solanaceae species found that NCBI annotations were generally longer and concluded that

“Overall, the large majority of FBP homologs detected from these various annotation sources were in close agreement, but when they differed, **homologs from the NCBI annotations were generally longer and more abundant than from the genome-specific databases. These results suggest that reannotation of genome sequences using a unified annotation strategy, such as employed by the NCBI pipeline, may be preferable for improved consistency during comparative genomics research.”**

*Li et. al. Genes (Basel). 2019 Jul 25;10(8). Assessing Anthocyanin Biosynthesis in Solanaceae as a Model Pathway for Secondary Metabolism.

Thank you.

RefSeq/Gene

Terence Murphy
Eric Cox
Catherine Farrell
Tamara Goldfarb
Diana Haddad
John Jackson
Vinita Joardar
Kelly McGarvey
Michael Murphy
Shashi Pujar
Bhanu Rajput
Sanjida Rangwala
Lillian Riddick
Barbara Robbertse
Brian Smith-White
Pooja Strope
David Webb

RefSeq Developers

Alex Astashyn
Olga Ermolaeva
Vamsi Kodali
Craig Wallin

Annotation Pipeline

Francoise Thibaud-Nissen
Paul Kitts
Mike Dicuccio
Wratko Hlavina
Avi Kimchi

Jinna Choi
Boris Kiryutin
Patrick Masterson
Eyal Mozes
Anton Perkov
Dan Rausch
Robert Smith
Alexandre Souvorov

GDV/Remap/GBench

Valerie Schneider
Peter Meric
Nathan Bouk
Hsiu-Chuan Chen
Cliff Clausen
Anatoliy Kuznetsov

A cast of thousands

Ken Katz
Michael Ovetsky
Lukas Wagner
Andrei Shkeda
Donna Maglott
Kim Pruitt
Jim Ostell

Watch NCBI News for updates!
http://www.ncbi.nlm.nih.gov/news/
https://www.youtube.com/user/NCBINLM

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

NCBI Genome Resources Workshop

Monday January 13, 2020, 12:50 – 3:00 pm, Pacific Salon 1

Time	Topic
12:55 – 1:15	NCBI Wants Your Sequence Data! How Do I Get It There? *Ilene Mizrachi*
1:15 – 1:35	Annotation of Eukaryote Genomes at NCBI *Jinna Hoffman*
1:35 – 1:55	Accessing Homologous Gene Datasets at NCBI *Nuala O'Leary*
1:55 – 2:15	The New PubMed Is Here! *Kathi Canese*
2:15 – 2:35	Taxonomy Lookup; Data Retrieval: How to Find and Stream Genomic Data in the Cloud! *Ben Busby*

Visit NCBI Booth **321** Contact us **info@ncbi.nlm.nih.gov**

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Watch NCBI News for updates!
http://www.ncbi.nlm.nih.gov/news/
https://www.youtube.com/user/NCBINLM

Also from NCBI!

Visit NCBI Booth **321** Contact us **info@ncbi.nlm.nih.gov**

Day	Time	Topic
Saturday	PENDING *Royal Palm Salon 3-4*	PENDING *Aquaculture*
Sunday	10:30 am *Town & Country*	**Stand-Alone PGAP: The NCBI Open-Source Pipeline for the Annotation of Prokaryotic Genomes** *Computational Gene Discovery*
Sunday	12:25 pm *San Diego*	**Genomic Resources for Agricultural Animals at NCBI** *Cattle/Sheep/Goat 2*
Sunday	1:42 pm *Pacific Salon 1*	**NCBI RefSeq Resources for Plant Genomics** *Functional Genomics*
Monday	12:50 pm – 3:00 pm *Pacific Salon 1*	**NCBI Genome Resources Workshop**
Tuesday	11:10 am *California*	**NCBI BLAST: Enhanced Web Usability through New Result Page and Effective Genomic Data Access** *Digital Tools and Resources Session 3*
Wednesday	11:50 am *California*	**Federated Cloud Access to Datasets through Indexing and/or Graphs!** *Digital Tools and Resources Session 4*